CALCULUS

Gregory Hartman et al.
Virginia Military Institute

LibreTexts™

Virginia Military Institute
Calculus

Gregory Hartman et al.

This text is disseminated via the Open Education Resource (OER) LibreTexts Project (https://LibreTexts.org) and like the thousands of other texts available within this powerful platform, it is freely available for reading, printing, and "consuming."

The LibreTexts mission is to bring together students, faculty, and scholars in a collaborative effort to provide an accessible, and comprehensive platform that empowers our community to develop, curate, adapt, and adopt openly licensed resources and technologies; through these efforts we can reduce the financial burden born from traditional educational resource costs, ensuring education is more accessible for students and communities worldwide.

Most, but not all, pages in the library have licenses that may allow individuals to make changes, save, and print this book. Carefully consult the applicable license(s) before pursuing such effects. Instructors can adopt existing LibreTexts texts or Remix them to quickly build course-specific resources to meet the needs of their students. Unlike traditional textbooks, LibreTexts' web based origins allow powerful integration of advanced features and new technologies to support learning.

LibreTexts is the adaptable, user-friendly non-profit open education resource platform that educators trust for creating, customizing, and sharing accessible, interactive textbooks, adaptive homework, and ancillary materials. We collaborate with individuals and organizations to champion open education initiatives, support institutional publishing programs, drive curriculum development projects, and more.

The LibreTexts libraries are Powered by NICE CXone Expert and was supported by the Department of Education Open Textbook Pilot Project, the California Education Learning Lab, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

Have questions or comments? For information about adoptions or adaptions contact info@LibreTexts.org or visit our main website at https://LibreTexts.org.

The LibreTexts name, logo, and book covers are trademarked by LibreTexts, Inc. (a 501(c)(3) not-for-profit organization) and protected against unauthorized use.

This text was compiled on 07/30/2026

- **InfoPage** by Delmar Larsen is licensed Public Domain.

TABLE OF CONTENTS

Licensing

1: Limits

- 1.1: An Introduction to Limits
- 1.2: Epsilon-Delta Definition of a Limit
- 1.3: Finding Limits Analytically
- 1.4: One Sided Limits
- 1.5: Continuity
- 1.6: Limits Involving Infinity
- 1.E: Applications of Limits (Exercises)

2: Derivatives

- 2.1: Instantaneous Rates of Change- The Derivative
- 2.2: Interpretations of the Derivative
- 2.3: Basic Differentiation Rules
- 2.4: The Product and Quotient Rules
- 2.5: The Chain Rule
- 2.6: Implicit Differentiation
- 2.7: Derivatives of Inverse Functions
- 2.E: Applications of Derivatives(Exercises)

3: The Graphical Behavior of Functions

- 3.1: Extreme Values
- 3.2: The Mean Value Theorem
- 3.3: Increasing and Decreasing Functions
- 3.4: Concavity and the Second Derivative
- 3.5: Curve Sketching
- 3.E: Applications of the Graphical Behavior of Functions(Exercises)

4: Applications of the Derivative

- 4.1: Newton's Method
- 4.2: Related Rates
- 4.3: Optimization
- 4.4: Differentials
- 4.E: Applications of Derivatives (Exercises)

5: Integration

- 5.1: Antiderivatives and Indefinite Integration
- 5.2: The Definite Integral
- 5.3: Riemann Sums
- 5.4: The Fundamental Theorem of Calculus
- 5.5: Numerical Integration
- 5.E: Applications of Integration (Exercises)

6: Techniques of Integration

- 6.1: Substitution
- 6.2: Integration by Parts
- 6.3: Trigonometric Integrals
- 6.4: Trigonometric Substitution
- 6.5: Partial Fraction Decomposition
- 6.6: Hyperbolic Functions
- 6.7: L'Hopital's Rule
- 6.8: Improper Integration
- 6.E: Applications of Antidifferentiation (Exercises)
- Index

7: Applications of Integration

- 7.1: Area Between Curves
- 7.2: Volume by Cross-Sectional Area- Disk and Washer Methods
- 7.3: The Shell Method
- 7.4: Arc Length and Surface Area
- 7.5: Work
- 7.6: Fluid Forces
- 7.E: Applications of Integration (Exercises)

8: Sequences and Series

- 8.1: Sequences
- 8.2: Infinite Series
- 8.3: Integral and Comparison Tests
- 8.4: Ratio and Root Tests
- 8.5: Alternating Series and Absolute Convergence
- 8.6: Power Series
- 8.7: Taylor Polynomials
- 8.8: Taylor Series
- 8.E: Applications of Sequences and Series (Exercises)

9: Curves in the Plane

- 9.1: Conic Sections
- 9.2: Parametric Equations
- 9.3: Calculus and Parametric Equations
- 9.4: Introduction to Polar Coordinates
- 9.5: Calculus and Polar Functions
- 9.E: Applications of Curves in a Plane (Exercises)

10: Vectors

- 10.1: Introduction to Cartesian Coordinates in Space
- 10.2: An Introduction to Vectors
- 10.3: The Dot Product
- 10.4: The Cross Product
- 10.5: Lines
- 10.6: Planes
- 10.E: Applications of Vectors (Exercises)

11: Vector-Valued Functions

- 11.1: Vector–Valued Functions
- 11.2: Calculus and Vector-Valued Functions
- 11.3: The Calculus of Motion
- 11.4: Unit Tangent and Normal Vectors
- 11.5: The Arc Length Parameter and Curvature
- 11.E: Applications of Vector Valued Functions (Exercises)

12: Functions of Several Variables

- 12.1: Introduction to Multivariable Functions
- 12.2: Limits and Continuity of Multivariable Functions
- 12.3: Partial Derivatives
- 12.4: Differentiability and the Total Differential
- 12.5: The Multivariable Chain Rule
- 12.6: Directional Derivatives
- 12.7: Tangent Lines, Normal Lines, and Tangent Planes
- 12.8: Extreme Values
- 12.E: Applications of Functions of Several Variables (Exercises)

13: Multiple Integration

- 13.1: Iterated Integrals and Area
- 13.2: Double Integration and Volume
- 13.3: Double Integration with Polar Coordinates
- 13.4: Center of Mass
- 13.5: Surface Area
- 13.6: Volume Between Surfaces and Triple Integration
- 13.E: Applications of Multiple Integration (Exercises)

14: Appendix

- 14.1: Section 1-
- 14.2: Section 2-
- 14.3: Section 3-
- 14.4: Section 4-
- 14.5: Section 5-
- 14.6: Section 6-

Index

Glossary

Detailed Licensing

Licensing

A detailed breakdown of this resource's licensing can be found in ***Back Matter/Detailed Licensing****.*

CHAPTER OVERVIEW

1: Limits

Calculus means "a method of calculation or reasoning." When one computes the sales tax on a purchase, one employs a simple calculus. When one finds the area of a polygonal shape by breaking it up into a set of triangles, one is using another calculus. Proving a theorem in geometry employs yet another calculus. Despite the wonderful advances in mathematics that had taken place into the first half of the 17th century, mathematicians and scientists were keenly aware of what *they could not do.* (This is true even today.) In particular, two important concepts eluded mastery by the great thinkers of that time: area and rates of change.

- Area seems innocuous enough; areas of circles, rectangles, parallelograms, etc., are standard topics of study for students today just as they were then. However, the areas of *arbitrary* shapes could not be computed, even if the boundary of the shape could be described exactly.
- Rates of change were also important. When an object moves at a constant rate of change, then "distance = rate × time." But what if the rate is not constant -- can distance still be computed? Or, if distance is known, can we discover the rate of change?

It turns out that these two concepts were related. Two mathematicians, Sir Isaac Newton and Gottfried Leibniz, are credited with independently formulating a system of computing that solved the above problems and showed how they were connected. Their system of reasoning was "a" calculus. However, as the power and importance of their discovery took hold, it became known to many as "the" calculus. Today, we generally shorten this to discuss "calculus." The foundation of "the calculus" is the *limit*. It is a tool to describe a particular behavior of a function. This chapter begins our study of the limit by approximating its value graphically and numerically. After a formal definition of the limit, properties are established that make "finding limits" tractable. Once the limit is understood, then the problems of area and rates of change can be approached.

1.1: An Introduction to Limits
1.2: Epsilon-Delta Definition of a Limit
1.3: Finding Limits Analytically
1.4: One Sided Limits
1.5: Continuity
1.6: Limits Involving Infinity
1.E: Applications of Limits (Exercises)

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 1: Limits is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

1.1: An Introduction to Limits

We begin our study of *limits* by considering examples that demonstrate key concepts that will be explained as we progress.

Consider the function $y = \frac{\sin x}{x}$. When x is near the value 1, what value (if any) is y near?

While our question is not precisely formed (what constitutes "near the value 1"?), the answer does not seem difficult to find. One might think first to look at a graph of this function to approximate the appropriate y values.

Consider Figure 1, where $y = \frac{\sin x}{x}$ is graphed. For values of x near 1, it seems that y takes on values near 0.85. In fact, when $x = 1$, then $y = \frac{\sin 1}{1} \approx 0.84$, so it makes sense that when x is "near" 1, y will be "near" 0.84.

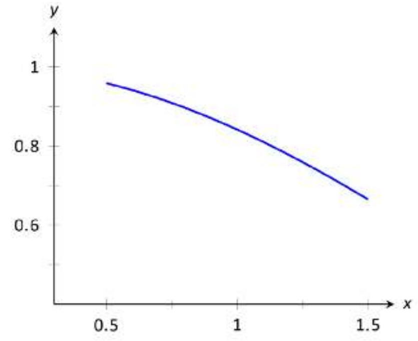


FIGURE 1.1:$\sin(x)/x$ near $x = 1$.

Consider this again at a different value for x. When x is near 0, what value (if any) is y near? By considering Figure 1.2, one can see that it seems that y takes on values near 1. But what happens when $x = 0$? We have

$$y \rightarrow \frac{\sin 0}{0} \rightarrow'' \frac{0}{0}''.$$

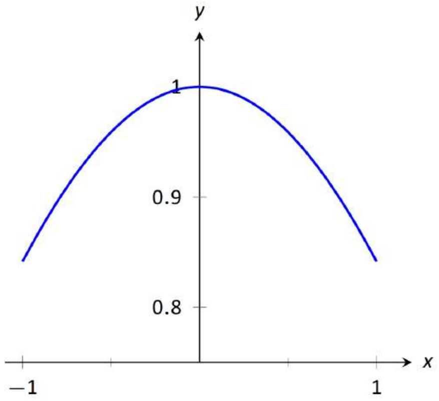


FIGURE 1.2:$\sin(x)/x$ near $x = 0$.

The expression "$0/0$" has no value; it is *indeterminate*. Such an expression gives no information about what is going on with the function nearby. We cannot find out how y behaves near $x = 0$ for this function simply by letting $x = 0$.

Finding a limit entails understanding how a function behaves near a particular value of x. Before continuing, it will be useful to establish some notation. Let $y = f(x)$; that is, let y be a function of x for some function f. The expression "the limit of y as x approaches 1" describes a number, often referred to as L, that y nears as x nears 1. We write all this as

$$\lim_{x\to 1} y = \lim_{x\to 1} f(x) = L.$$

This is not a complete definition (that will come in the next section); this is a pseudo-definition that will allow us to explore the idea of a limit.

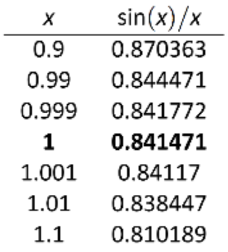
Above, where $f(x) = \sin(x)/x$, we approximated

$$\lim_{x\to 1} \frac{\sin x}{x} \approx 0.84 \quad \text{and} \quad \lim_{x\to 0} \frac{\sin x}{x} \approx 1.$$

(We *approximated* these limits, hence used the "≈" symbol, since we are working with the pseudo-definition of a limit, not the actual definition.)

Once we have the true definition of a limit, we will find limits *analytically*; that is, exactly using a variety of mathematical tools. For now, we will *approximate* limits both graphically and numerically. Graphing a function can provide a good approximation, though often not very precise. Numerical methods can provide a more accurate approximation. We have already approximated limits graphically, so we now turn our attention to numerical approximations.

Consider again $\lim_{x\to 1} \sin(x)/x$. To approximate this limit numerically, we can create a table of x and $f(x)$ values where x is "near" 1. This is done in Figure 1.3.

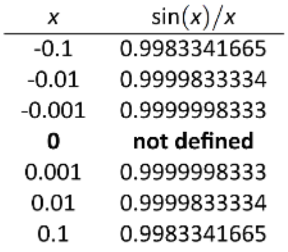


x	sin(x)/x
0.9	0.870363
0.99	0.844471
0.999	0.841772
1	**0.841471**
1.001	0.84117
1.01	0.838447
1.1	0.810189

FIGURE 1.3: Values of $\sin(x)/x$ with x near 1.

Notice that for values of x near 1, we have $\sin(x)/x$ near 0.841. The $x = 1$ row is in bold to highlight the fact that when considering limits, we are *not* concerned with the value of the function at that particular x value; we are only concerned with the values of the function when x is *near* 1.

Now approximate $\lim_{x\to 0} \sin(x)/x$ numerically. We already approximated the value of this limit as 1 graphically in Figure 1.2. The table in Figure 1.4 shows the value of $\sin(x)/x$ for values of x near 0. Ten places after the decimal point are shown to highlight how close to 1 the value of $\sin(x)/x$ gets as x takes on values very near 0. We include the $x = 0$ row in bold again to stress that we are not concerned with the value of our function at $x = 0$, only on the behavior of the function *near* 0.



x	sin(x)/x
-0.1	0.9983341665
-0.01	0.9999833334
-0.001	0.9999998333
0	**not defined**
0.001	0.9999998333
0.01	0.9999833334
0.1	0.9983341665

FIGURE 1.4: Values of $\sin(x)/x$ with x near 1.

This numerical method gives confidence to say that 1 is a good approximation of $\lim_{x\to 0} \sin(x)/x$; that is,

$$\lim_{x\to 0} \sin(x)/x \approx 1. \tag{1.1.1}$$

Later we will be able to prove that the limit is *exactly* 1.

We now consider several examples that allow us explore different aspects of the limit concept.

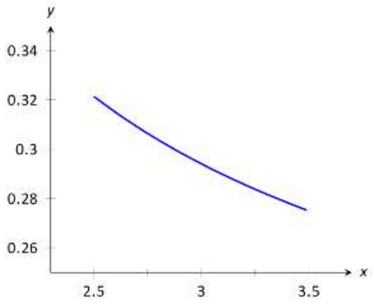
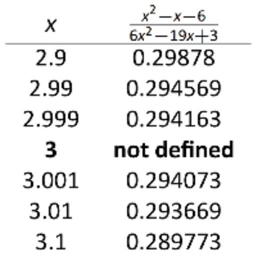


FIGURE 1.5: Graphically approximating a limit in Example 1.



x	$\frac{x^2-x-6}{6x^2-19x+3}$
2.9	0.29878
2.99	0.294569
2.999	0.294163
3	**not defined**
3.001	0.294073
3.01	0.293669
3.1	0.289773

FIGURE 1.6: Numerically approximating a limit in Example 1.

Example 1: Approximating the value of a limit

Use graphical and numerical methods to approximate

$$\lim_{x\to 3} \frac{x^2-x-6}{6x^2-19x+3}. \tag{1.1.2}$$

Solution:
To graphically approximate the limit, graph

$$y=(x^2-x-6)/(6x^2-19x+3) \tag{1.1.3}$$

on a small interval that contains 3. To numerically approximate the limit, create a table of values where the x values are near 3. This is done in Figures 1.5 and 1.6, respectively.

The graph shows that when x is near 3, the value of y is very near 0.3. By considering values of x near 3, we see that $y=0.294$ is a better approximation. The graph and the table imply that

$$\lim_{x\to 3} \frac{x^2-x-6}{6x^2-19x+3} \approx 0.294. \tag{1.1.4}$$

This example may bring up a few questions about approximating limits (and the nature of limits themselves).

1. If a graph does not produce as good an approximation as a table, why bother with it?
2. How many values of x in a table are "enough?" In the previous example, could we have just used $x=3.001$ and found a fine approximation?

Graphs are useful since they give a visual understanding concerning the behavior of a function. Sometimes a function may act "erratically" near certain x values which is hard to discern numerically but very plain graphically. Since graphing utilities are very accessible, it makes sense to make proper use of them.

Since tables and graphs are used only to *approximate* the value of a limit, there is not a firm answer to how many data points are "enough." Include enough so that a trend is clear, and use values (when possible) both less than and greater than the value in question. In Example 1, we used both values less than and greater than 3. Had we used just $x=3.001$, we might have been tempted to conclude that the limit had a value of 0.3. While this is not far off, we could do better. Using values "on both sides of 3" helps us identify trends.

Example 2: Approximating the value of a limit

Graphically and numerically approximate the limit of $f(x)$ as x approaches 0, where

$$f(x)=\begin{cases} x+1 & x<0 \\ -x^2+1 & x>0 \end{cases} \tag{1.1.5}$$

Solution:
Again we graph $f(x)$ and create a table of its values near $x=0$ to approximate the limit. Note that this is a piecewise defined function, so it behaves differently on either side of 0. Figure 1.7 shows a graph of $f(x)$, and on either side of 0 it seems the y values approach 1. Note that $f(0)$ is not actually defined, as indicated in the graph with the open circle.

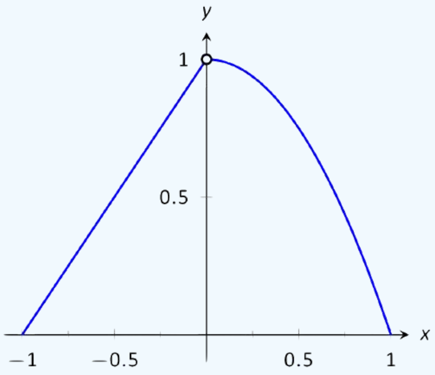


FIGURE 1.7: Graphically approximating a limit in Example 2.

x	$f(x)$
-0.1	0.9
-0.01	0.99
-0.001	0.999
0.001	0.999999
0.01	0.9999
0.1	0.99

FIGURE 1.8: Numerically approximating a limit in Example 2.

The table shown in Figure 1.8 shows values of $f(x)$ for values of x near 0. It is clear that as x takes on values very near 0, $f(x)$ takes on values very near 1. It turns out that if we let $x=0$ for either "piece" of $f(x)$, 1 is returned; this is significant and we'll return to this idea later.

The graph and table allow us to say that $\lim_{x\to 0} f(x) \approx 1$; in fact, we are probably very sure it *equals* 1.

Identifying When Limits Do Not Exist

A function may not have a limit for all values of x. That is, we cannot say $\lim_{x\to c} f(x) = L$ for some numbers L for all values of c, for there may not be a number that $f(x)$ is approaching. There are three ways in which a limit may fail to exist.

1. The function $f(x)$ may approach different values on either side of c.
2. The function may grow without upper or lower bound as x approaches c.
3. The function may oscillate as x approaches c.

We'll explore each of these in turn.

Example 3: Different Values Approached From Left and Right

Explore why $\lim_{x\to 1} f(x)$ does not exist, where

$$f(x) = \begin{cases} x^2-2x+3 & x \le 1 \\ x & x > 1 \end{cases} \tag{1.1.6}$$

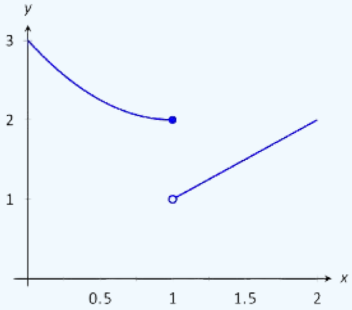
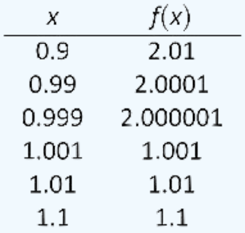


FIGURE 1.9: Observing no limit as $x \to 1$ in Example 3.



x	$f(x)$
0.9	2.01
0.99	2.0001
0.999	2.000001
1.001	1.001
1.01	1.01
1.1	1.1

FIGURE 1.10: Values of $f(x)$ near $x = 1$ in Example 3.

Solution:
A graph of $f(x)$ around $x = 1$ and a table are given Figures 1.9 and 1.10, respectively. It is clear that as x approaches 1, $f(x)$ does not seem to approach a single number. Instead, it seems as though $f(x)$ approaches two different numbers. When considering values of x less than 1 (approaching 1 from the left), it seems that $f(x)$ is approaching 2; when considering values of x greater than 1 (approaching 1 from the right), it seems that $f(x)$ is approaching 1. Recognizing this behavior is important; we'll study this in greater depth later. Right now, it suffices to say that the limit does not exist since $f(x)$ is not approaching one value as x approaches 1.

Example 4: The Function Grows Without Bound

Explore why $\lim_{x\to 1} 1/(x-1)^2$ does not exist.

Solution:
A graph and table of $f(x) = 1/(x-1)^2$ are given in Figures 1.11 and 1.12, respectively. Both show that as x approaches 1, $f(x)$ grows larger and larger.

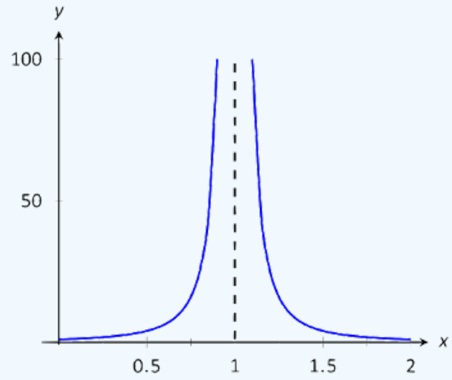
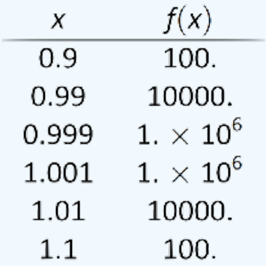


FIGURE 1.11: Observing no limit as $x \to 1$ in Example 4.



x	$f(x)$
0.9	100.
0.99	10000.
0.999	$1. \times 10^6$
1.001	$1. \times 10^6$
1.01	10000.
1.1	100.

FIGURE 1.12: Values of $f(x)$ near $x = 1$ in Example 4.

We can deduce this on our own, without the aid of the graph and table. If x is near 1, then $(x-1)^2$ is very small, and:

$$\frac{1}{\text{very small number}} = \text{very large number.} \tag{1.1.7}$$

Since $f(x)$ is not approaching a single number, we conclude that

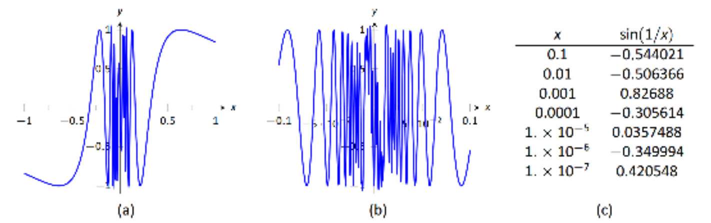
$$\lim_{x\to 1} \frac{1}{(x-1)^2} \tag{1.1.8}$$

does not exist.

Example 5: The Function Oscillates

Explore why $\lim_{x\to 0}\sin(1/x)$ does not exist.

Solution:
Two graphs of $f(x)=\sin(1/x)$ are given in Figures 1.13. Figure 1.13(a) shows $f(x)$ on the interval $[-1,1]$; notice how $f(x)$ seems to oscillate near $x=0$. One might think that despite the oscillation, as x approaches 0, $f(x)$ approaches 0. However, Figure 1.13(b) zooms in on $\sin(1/x)$, on the interval $[-0.1,0.1]$. Here the oscillation is even more pronounced. Finally, in the table in Figure 1.13(c), we see $\sin(x)/x$ evaluated for values of x near 0. As x approaches 0, $f(x)$ does not appear to approach any value.



x	$\sin(1/x)$
0.1	−0.544021
0.01	−0.506366
0.001	0.82688
0.0001	−0.305614
$1. \times 10^{-5}$	0.0357488
$1. \times 10^{-6}$	−0.349994
$1. \times 10^{-7}$	0.420548

FIGURE 1.13: Observing that $f(x)=\sin(1/x)$ has no limit as $x\to 0$ in Example 5.

It can be shown that in reality, as x approaches 0, $\sin(1/x)$ takes on all values between -1 and 1 infinite times! Because of this oscillation,

$\lim_{x\to 0}\sin(1/x)$ does not exist.

Limits of Difference Quotients

We have approximated limits of functions as x approached a particular number. We will consider another important kind of limit after explaining a few key ideas.

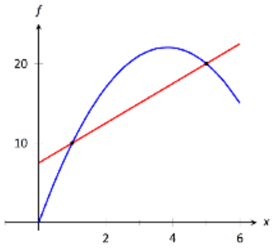


FIGURE 1.14: Interpreting a difference quotient as the slope of a secant line.

Let $f(x)$ represent the position function, in feet, of some particle that is moving in a straight line, where x is measured in seconds. Let's say that when $x=1$, the particle is at position 10 ft., and when $x=5$, the particle is at 20 ft. Another way of expressing this is to say

$$f(1)=10 \quad \text{and} \quad f(5)=20. \tag{1.1.9}$$

Since the particle traveled 10 feet in 4 seconds, we can say the particle's *average velocity* was 2.5 ft/s. We write this calculation using a "quotient of differences," or, a *difference quotient*:

$$\frac{f(5)-f(1)}{5-1}=\frac{10}{4}=2.5\text{ft/s}. \tag{1.1.10}$$

This difference quotient can be thought of as the familiar "rise over run" used to compute the slopes of lines. In fact, that is essentially what we are doing: given two points on the graph of f, we are finding the slope of the *secant line* through those two points. See Figure 1.14.

Now consider finding the average speed on another time interval. We again start at $x=1$, but consider the position of the particle h seconds later. That is, consider the positions of the particle when $x=1$ and when $x=1+h$. The difference quotient is now

$$\frac{f(1+h)-f(1)}{(1+h)-1}=\frac{f(1+h)-f(1)}{h}. \tag{1.1.11}$$

Let $f(x)=-1.5x^2+11.5x$; note that $f(1)=10$ and $f(5)=20$, as in our discussion. We can compute this difference quotient for all values of h (even negative values!) except $h=0$, for then we get "0/0," the indeterminate form introduced earlier. For all values $h \neq 0$, the difference quotient computes the average velocity of the particle over an interval of time of length h starting at $x=1$.

For small values of h, i.e., values of h close to 0, we get average velocities over very short time periods and compute secant lines over small intervals. See Figure 1.15. This leads us to wonder what the limit of the difference quotient is as h approaches 0. That is,

$$\lim_{h\to 0}\frac{f(1+h)-f(1)}{h}=\ ? \tag{1.1.12}$$

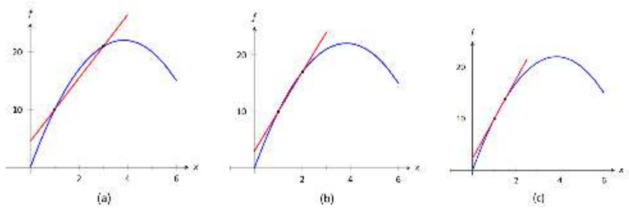
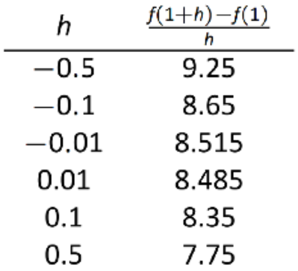


FIGURE 1.15: Secant lines of $f(x)$ at $x=1$ and $x=1+h$, for shrinking values of h (i.e., $h\to 0$).

As we do not yet have a true definition of a limit nor an exact method for computing it, we settle for approximating the value. While we could graph the difference quotient (where the x-axis would represent h values and the y-axis would represent values of the difference quotient) we settle for making a table. See Figure 1.16. The table gives us reason to assume the value of the limit is about 8.5.



h	$\frac{f(1+h)-f(1)}{h}$
-0.5	9.25
-0.1	8.65
-0.01	8.515
0.01	8.485
0.1	8.35
0.5	7.75

FIGURE 1.16: The difference quotient evaluated at values of h near 0.

Proper understanding of limits is key to understanding calculus. With limits, we can accomplish seemingly impossible mathematical things, like adding up an infinite number of numbers (and not get infinity) and finding the slope of a line between two points, where the "two points" are actually the same point. These are not just mathematical curiosities; they allow us to link position, velocity and acceleration together, connect cross-sectional areas to volume, find the work done by a variable force, and much more.

In the next section we give the formal definition of the limit and begin our study of finding limits analytically. In the following exercises, we continue our introduction and approximate the value of limits.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 1.1: An Introduction to Limits is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

1.2: Epsilon-Delta Definition of a Limit

This section introduces the formal definition of a limit. Many refer to this as "the epsilon--delta," definition, referring to the letters ϵ and δ of the Greek alphabet.

Before we give the actual definition, let's consider a few informal ways of describing a limit. Given a function $y = f(x)$ and an x-value, c, we say that "the limit of the function f, as x approaches c, is a value L":

1. if "y tends to L" as "x tends to c."
2. if "y approaches L" as "x approaches c."
3. if "y is near L" whenever "x is near c."

The problem with these definitions is that the words "tends," "approach," and especially "near" are not exact. In what way does the variable x tend to, or approach, c? How near do x and y have to be to c and L, respectively?

The definition we describe in this section comes from formalizing **3**. A quick restatement gets us closer to what we want:

$\mathbf{3}'$. If x is within a certain *tolerance level* of c, then the corresponding value $y = f(x)$ is within a certain *tolerance level* of L.

The traditional notation for the x-tolerance is the lowercase Greek letter delta, or δ, and the y-tolerance is denoted by lowercase epsilon, or ϵ. One more rephrasing of $\mathbf{3}'$ nearly gets us to the actual definition:

$\mathbf{3}''$. If x is within δ units of c, then the corresponding value of y is within ϵ units of L.

We can write "x is within δ units of c" mathematically as

$$|x - c| < \delta, \qquad \text{which is equivalent to} \qquad c - \delta < x < c + \delta. \tag{1.2.1}$$

Letting the symbol "$\longrightarrow$" represent the word "implies," we can rewrite $\mathbf{3}''$ as

$$|x - c| < \delta \longrightarrow |y - L| < \epsilon \qquad \text{or} \qquad c - \delta < x < c + \delta \longrightarrow L - \epsilon < y < L + \epsilon. \tag{1.2.2}$$

The point is that δ and ϵ, being tolerances, can be any positive (but typically small) values. Finally, we have the formal definition of the limit with the notation seen in the previous section.

Definition 1: The Limit of a Function f

Let I be an open interval containing c, and let f be a function defined on I, except possibly at c. The **limit of** $f(x)$, **as** x **approaches** c, **is** L, denoted by

$$\lim_{x \to c} f(x) = L, \tag{1.2.3}$$

means that given any $\epsilon > 0$, there exists $\delta > 0$ such that for all $x \neq c$, if $|x - c| < \delta$, then $|f(x) - L| < \epsilon$.

(Mathematicians often enjoy writing ideas without using any words. Here is the wordless definition of the limit:

$$\lim_{x \to c} f(x) = L \iff \forall\, \epsilon > 0, \exists\, \delta > 0 \; s.t. \; 0 < |x - c| < \delta \longrightarrow |f(x) - L| < \epsilon.) \tag{1.2.4}$$

Note the order in which ϵ and δ are given. In the definition, the y-tolerance ϵ is given *first* and then the limit will exist **if** we can find an x-tolerance δ that works.

An example will help us understand this definition. Note that the explanation is long, but it will take one through all steps necessary to understand the ideas.

Example 6: Evaluating a limit using the definition

Show that $\lim_{x \to 4} \sqrt{x} = 2$.

Solution:
Before we use the formal definition, let's try some numerical tolerances. What if the y tolerance is 0.5, or $\epsilon = 0.5$? How close to 4 does x have to be so that y is within 0.5 units of 2, i.e., $1.5 < y < 2.5$? In this case, we can proceed as follows:

$$1.5 < y < 2.5 \tag{1.2.5}$$

$$1.5 < \sqrt{x} < 2.5 \tag{1.2.6}$$

$$1.5^2 < x < 2.5^2 \tag{1.2.7}$$

$$1.5^2 < x < 2.5^2 \tag{1.2.7}$$

$$2.25 < x < 6.25. \tag{1.2.8}$$

So, what is the desired x tolerance? Remember, we want to find a symmetric interval of x values, namely $4-\delta < x < 4+\delta$. The lower bound of 2.25 is 1.75 units from 4; the upper bound of 6.25 is 2.25 units from 4. We need the smaller of these two distances; we must have $\delta \leq 1.75$. See Figure 1.17.

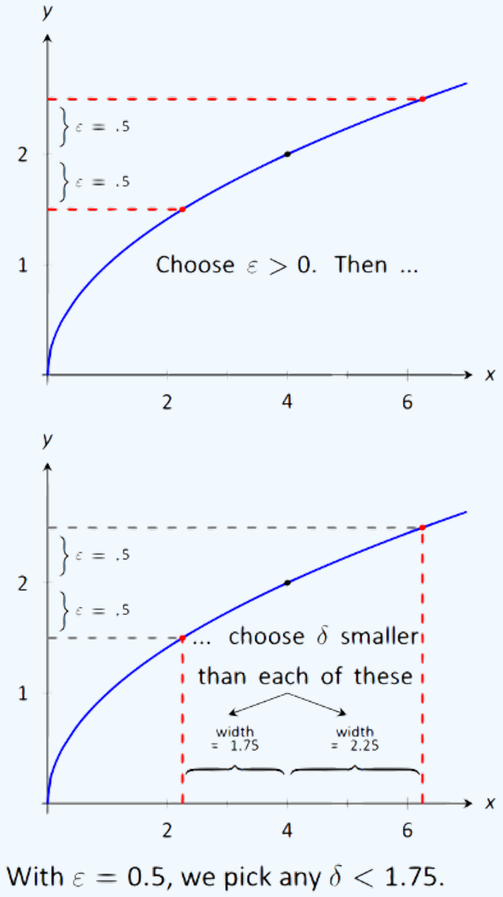


FIGURE 1.17: Illustrating the $\epsilon - \delta$ process.

Given the y tolerance $\epsilon = 0.5$, we have found an x tolerance, $\delta \leq 1.75$, such that whenever x is within δ units of 4, then y is within ϵ units of 2. That's what we were trying to find.

Let's try another value of ϵ.

What if the y tolerance is 0.01, i.e., $\epsilon = 0.01$? How close to 4 does x have to be in order for y to be within 0.01 units of 2 (or $1.99 < y < 2.01$)? Again, we just square these values to get $1.99^2 < x < 2.01^2$, or

$$3.9601 < x < 4.0401. \tag{1.2.9}$$

What is the desired x tolerance? In this case we must have $\delta \leq 0.0399$, which is the minimum distance from 4 of the two bounds given above.

Note that in some sense, it looks like there are two tolerances (below 4 of 0.0399 units and above 4 of 0.0401 units). However, we couldn't use the larger value of 0.0401 for δ since then the interval for x would be $3.9599 < x < 4.0401$ resulting in y values of $1.98995 < y < 2.01$ (which contains values NOT within 0.01 units of 2).

What we have so far: if $\epsilon = 0.5$, then $\delta \leq 1.75$ and if $\epsilon = 0.01$, then $\delta \leq 0.0399$. A pattern is not easy to see, so we switch to general ϵ try to determine δ symbolically. We start by assuming $y = \sqrt{x}$ is within ϵ units of 2:

$$\begin{aligned} |y-2| &< \epsilon & \\ -\epsilon < y-2 &< \epsilon & \text{(Definition of absolute value)} \\ -\epsilon < \sqrt{x}-2 &< \epsilon & (y=\sqrt{x}) \\ 2-\epsilon < \sqrt{x} &< 2+\epsilon & \text{(Add 2)} \\ (2-\epsilon)^2 < x &< (2+\epsilon)^2 & \text{(Square all)} \\ 4-4\epsilon+\epsilon^2 < x &< 4+4\epsilon+\epsilon^2 & \text{(Expand)} \\ 4-(4\epsilon-\epsilon^2) < x &< 4+(4\epsilon+\epsilon^2). & \text{(Rewrite in the desired form)} \end{aligned}$$

The "desired form" in the last step is "$4 - \mathit{something} < x < 4 + \mathit{something}$." Since we want this last interval to describe an x tolerance around 4, we have that either $\delta \le 4\epsilon - \epsilon^2$ or $\delta \le 4\epsilon + \epsilon^2$, whichever is smaller:

$$\delta \le \min\{4\epsilon-\epsilon^2, 4\epsilon+\epsilon^2\}. \tag{1.2.10}$$

Since $\epsilon > 0$, the minimum is $\delta \le 4\epsilon - \epsilon^2$. That's the formula: given an ϵ, set $\delta \le 4\epsilon - \epsilon^2$.

We can check this for our previous values. If $\epsilon = 0.5$, the formula gives $\delta \le 4(0.5) - (0.5)^2 = 1.75$ and when $\epsilon = 0.01$, the formula gives $\delta \le 4(0.01) - (0.01)^2 = 0.399$.

So given any $\epsilon > 0$, set $\delta \le 4\epsilon - \epsilon^2$. Then if $|x-4| < \delta$ (and $x \ne 4$), then $|f(x)-2| < \epsilon$, satisfying the definition of the limit. We have shown formally (and finally!) that $\lim_{x\to 4}\sqrt{x} = 2$.

> *Actually, it is a pain, but this won't work if $\epsilon \ge 4$. This shouldn't really occur since ϵ is supposed to be small, but it could happen. In the cases where $\epsilon \ge 4$, just take $\delta = 1$ and you'll be fine.*

The previous example was a little long in that we sampled a few specific cases of ϵ before handling the general case. Normally this is not done. The previous example is also a bit unsatisfying in that $\sqrt{4} = 2$; why work so hard to prove something so obvious? Many ϵ-δ proofs are long and difficult to do. In this section, we will focus on examples where the answer is, frankly, obvious, because the non--obvious examples are even harder. In the next section we will learn some theorems that allow us to evaluate limits *analytically*, that is, without using the ϵ-δ definition.

That is why theorems about limits are so useful! After doing a few more ϵ-δ proofs, you will really appreciate the analytical "short cuts" found in the next section.

Example 7: Evaluating a limit using the definition

Show that $\lim_{x\to 2} x^2 = 4$.

Solution

Let's do this example symbolically from the start. Let $\epsilon > 0$ be given; we want $|y-4| < \epsilon$, i.e., $|x^2-4| < \epsilon$. How do we find δ such that when $|x-2| < \delta$, we are guaranteed that $|x^2-4| < \epsilon$?

This is a bit trickier than the previous example, but let's start by noticing that $|x^2-4| = |x-2|\cdot|x+2|$. Consider:

$$|x^2-4| < \epsilon \longrightarrow |x-2|\cdot|x+2| < \epsilon \longrightarrow |x-2| < \frac{\epsilon}{|x+2|}. \tag{1.1}$$

Could we not set $\delta = \frac{\epsilon}{|x+2|}$?

We are close to an answer, but the catch is that δ must be a *constant* value (so it can't contain x). There is a way to work around this, but we do have to make an assumption. Remember that ϵ is supposed to be a small number, which implies that δ will also be a small value. In particular, we can (probably) assume that $\delta < 1$. If this is true, then $|x-2| < \delta$ would imply that $|x-2| < 1$, giving $1 < x < 3$.

Now, back to the fraction $\frac{\epsilon}{|x+2|}$. If $1 < x < 3$, then $3 < x+2 < 5$ (add 2 to all terms in the inequality). Taking reciprocals, we have

$$\frac{1}{5} < \frac{1}{|x+2|} < \frac{1}{3} \qquad \text{which implies} \tag{1.2.11}$$

$$\frac{1}{5} < \frac{1}{|x+2|} \qquad \text{which implies} \tag{1.2.12}$$

$$\frac{}{5} < \frac{}{|x+2|}. \qquad (1.2)$$

This suggests that we set $\delta \leq \frac{\epsilon}{5}$. To see why, let consider what follows when we assume $|x-2| < \delta$:

$$\begin{aligned}
|x-2| &< \delta \\
|x-2| &< \frac{\epsilon}{5} && \text{(Our choice of } \delta\text{)} \\
|x-2| \cdot |x+2| &< |x+2| \cdot \frac{\epsilon}{5} && \text{(Multiply by } |x+2|\text{)} \\
|x^2-4| &< |x+2| \cdot \frac{\epsilon}{5} && \text{(Combine left side)} \\
|x^2-4| &< |x+2| \cdot \frac{\epsilon}{5} < |x+2| \cdot \frac{\epsilon}{|x+2|} = \epsilon && \text{(Using (1.2) as long as } \delta < 1\text{)}
\end{aligned}$$

We have arrived at $|x^2-4| < \epsilon$ as desired. Note again, in order to make this happen we needed δ to first be less than 1. That is a safe assumption; we want ϵ to be arbitrarily small, forcing δ to also be small.

We have also picked δ to be smaller than "necessary." We could get by with a slightly larger δ, as shown in Figure 1.18. The dashed outer lines show the boundaries defined by our choice of ϵ. The dotted inner lines show the boundaries defined by setting $\delta = \epsilon/5$. Note how these dotted lines are within the dashed lines. That is perfectly fine; by choosing x within the dotted lines we are guaranteed that $f(x)$ will be within ϵ of 4.%If the value we eventually used for δ, namely $\epsilon/5$, is not less than 1, this proof won't work. For the final fix, we instead set δ to be the minimum of 1 and $\epsilon/5$. This way all calculations above work.

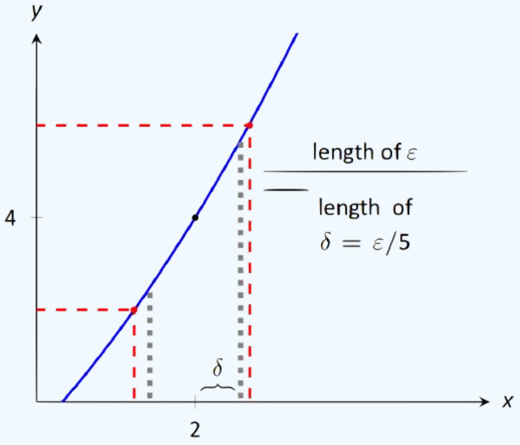


FIGURE 1.18: Choosing $\delta = \epsilon/5$ in Example 7.

In summary, given $\epsilon > 0$, set $\delta = \leq \epsilon/5$. Then $|x-2| < \delta$ implies $|x^2-4| < \epsilon$ (i.e. $|y-4| < \epsilon$) as desired. This shows that $\lim_{x\to 2} x^2 = 4$. Figure 1.18 gives a visualization of this; by restricting x to values within $\delta = \epsilon/5$ of 2, we see that $f(x)$ is within ϵ of 4.

Make note of the general pattern exhibited in these last two examples. In some sense, each starts out "backwards." That is, while we want to

1. start with $|x-c| < \delta$ and conclude that
2. $|f(x)-L| < \epsilon$,

we actually start by assuming

1. $|f(x)-L| < \epsilon$, then perform some algebraic manipulations to give an inequality of the form
2. $|x-c| <$ *something*.

When we have properly done this, the *something* on the "greater than" side of the inequality becomes our δ. We can refer to this as the "scratch--work" phase of our proof. Once we have δ, we can formally start with $|x-c| < \delta$ and use algebraic manipulations to conclude that $|f(x)-L| < \epsilon$, usually by using the same steps of our "scratch--work" in reverse order.

We highlight this process in the following example.

Example 8: Evaluating a limit using the definition

Prove that $\lim\limits_{x\to 1} x^3-2x=-1$.

Solution

We start our scratch--work by considering $|f(x)-(-1)|<\epsilon$:

$$\begin{aligned} |f(x)-(-1)| &< \epsilon && && (1.2.13)\\ |x^3-2x+1| &< \epsilon && \text{(Now factor)} && (1.2.14)\\ |(x-1)(x^2+x-1)| &< \epsilon && && (1.2.15)\\ |x-1| &< \frac{\epsilon}{|x^2+x-1|}. && && (1.3) \end{aligned}$$

We are at the phase of saying that $|x-1|<$ *something*, where *something* $=\epsilon/|x^2+x-1|$. We want to turn that *something* into δ.

Since x is approaching 1, we are safe to assume that x is between 0 and 2. So

$$\begin{aligned} 0<x&<2 \\ 0<x^2&<4. && \text{(squared each term)} \end{aligned}$$

Since $0<x<2$, we can add 0, x and 2, respectively, to each part of the inequality and maintain the inequality.

$$\begin{aligned} 0<x^2+x&<6 \\ -1<x^2+x-1&<5. && \text{(subtracted 1 from each part)} \end{aligned}$$

In Equation (1.3), we wanted $|x-1|<\epsilon/|x^2+x-1|$. The above shows that given any x in $[0,2]$, we know that

$$\begin{aligned} x^2+x-1&<5 && \text{which implies that}\\ \frac{1}{5}&<\frac{1}{x^2+x-1} && \text{which implies that}\\ \frac{\epsilon}{5}&<\frac{\epsilon}{x^2+x-1}. && && (1.4) \end{aligned}$$

So we set $\delta\le\epsilon/5$. This ends our scratch--work, and we begin the formal proof (which also helps us understand why this was a good choice of δ).

Given ϵ, let $\delta\le\epsilon/5$. We want to show that when $|x-1|<\delta$, then $|(x^3-2x)-(-1)|<\epsilon$. We start with $|x-1|<\delta$:

$$\begin{aligned} |x-1|&<\delta\\ |x-1|&<\frac{\epsilon}{5}\\ |x-1|&<\frac{\epsilon}{5}<\frac{\epsilon}{|x^2+x-1|} && \text{(for } x \text{ near 1, from Equation (1.4))}\\ |x-1|\cdot|x^2+x-1|&<\epsilon\\ |x^3-2x+1|&<\epsilon\\ |(x^3-2x)-(-1)|&<\epsilon, \end{aligned}$$

which is what we wanted to show. Thus $\lim\limits_{x\to 1} x^3-2x=-1$.

We illustrate evaluating limits once more.

Example 9: Evaluating a limit using the definition

Prove that $\lim\limits_{x\to 0} e^x=1$.

Solution

Symbolically, we want to take the equation $|e^x-1|<\epsilon$ and unravel it to the form $|x-0|<\delta$. Here is our scratch--work:

$$\begin{aligned} |e^x-1|&<\epsilon\\ -\epsilon<e^x-1&<\epsilon && \text{(Definition of absolute value)}\\ 1-\epsilon<e^x&<1+\epsilon && \text{(Add 1)}\\ \ln(1-\epsilon)<x&<\ln(1+\epsilon) && \text{(Take natural logs)} \end{aligned}$$

Making the safe assumption that $\epsilon < 1$ ensures the last inequality is valid (i.e., so that $\ln(1-\epsilon)$ is defined). We can then set δ to be the minimum of $|\ln(1-\epsilon)|$ and $\ln(1+\epsilon)$; i.e.,

$$\delta = \min\{|\ln(1-\epsilon)|, \ln(1+\epsilon)\} = \ln(1+\epsilon). \tag{1.2.16}$$

> *Recall* $\ln 1 = 0$ *and* $\ln x < 0$ *when* $0 < x < 1$. *So* $\ln(1-\epsilon) < 0$, *hence we consider its absolute value.*

Now, we work through the actual the proof:

$$\begin{aligned} |x-0| &< \delta \\ -\delta < x &< \delta & \text{(Definition of absolute value)} \\ -\ln(1+\epsilon) < x &< \ln(1+\epsilon). \\ \ln(1-\epsilon) < x &< \ln(1+\epsilon). & \text{(since } \ln(1-\epsilon) < -\ln(1+\epsilon)) \end{aligned}$$

The above line is true by our choice of δ and by the fact that since $|\ln(1-\epsilon)| > \ln(1+\epsilon)$ and $\ln(1-\epsilon) < 0$, we know $\ln(1-\epsilon) < -\ln(1+\epsilon)$.

$$\begin{aligned} 1-\epsilon < e^x &< 1+\epsilon & \text{(Exponentiate)} \\ -\epsilon < e^x - 1 &< \epsilon & \text{(Subtract 1)} \end{aligned}$$

In summary, given $\epsilon > 0$, let $\delta = \ln(1+\epsilon)$. Then $|x-0| < \delta$ implies $|e^x - 1| < \epsilon$ as desired. We have shown that $\lim\limits_{x\to 0} e^x = 1$.

We note that we could actually show that $\lim_{x\to c} e^x = e^c$ for any constant c. We do this by factoring out e^c from both sides, leaving us to show $\lim_{x\to c} e^{x-c} = 1$ instead. By using the substitution $u = x - c$, this reduces to showing $\lim_{u\to 0} e^u = 1$ which we just did in the last example. As an added benefit, this shows that in fact the function $f(x) = e^x$ is *continuous* at all values of x, an important concept we will define in Section 1.5.

This formal definition of the limit is not an easy concept grasp. Our examples are actually "easy" examples, using "simple" functions like polynomials, square--roots and exponentials. It is very difficult to prove, using the techniques given above, that $\lim\limits_{x\to 0} (\sin x)/x = 1$, as we approximated in the previous section.

There is hope. The next section shows how one can evaluate complicated limits using certain basic limits as building blocks. While limits are an incredibly important part of calculus (and hence much of higher mathematics), rarely are limits evaluated using the definition. Rather, the techniques of the following section are employed.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 1.2: Epsilon-Delta Definition of a Limit is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

1.3: Finding Limits Analytically

In Section 1.1 we explored the concept of the limit without a strict definition, meaning we could only make approximations. In the previous section we gave the definition of the limit and demonstrated how to use it to verify our approximations were correct. Thus far, our method of finding a limit is 1) make a really good approximation either graphically or numerically, and 2) verify our approximation is correct using a ϵ-δ proof.

> This process has its shortcomings, not the least of which is the fact that ϵ--δ proofs are cumbersome. This section gives a series of theorems which allow us to find limits much more quickly and intuitively.

Recognizing that ϵ-δ proofs are cumbersome, this section gives a series of theorems which allow us to find limits much more quickly and intuitively.

Suppose that $\lim\limits_{x\to 2} f(x)=2$ and $\lim\limits_{x\to 2} g(x)=3$.What is $\lim\limits_{x\to 2}(f(x)+g(x))$? Intuition tells us that the limit should be 5, as we expect limits to behave in a nice way. The following theorem states that already established limits do behave nicely.

Theorem 1.3.1: Basic Limit Properties

Let b, c, L and K be real numbers, let n be a positive integer, and let f and g be functions with the following limits:

$$\lim_{x\to c} f(x)=L \text{ and } \lim_{x\to c} g(x)=K.$$

The following limits hold.

1. Constants: $\lim\limits_{x\to c} b=b$
2. Identity: $\lim\limits_{x\to c} x=c$
3. Sums/Differences: $\lim\limits_{x\to c}(f(x)\pm g(x))=L\pm K$
4. Scalar Multiples: $\lim\limits_{x\to c} b\cdot f(x)=bL$
5. Products: $\lim\limits_{x\to c} f(x)\cdot g(x)=LK$
6. Quotients: $\lim\limits_{x\to c} f(x)/g(x)=L/K$, $(K\neq 0)$
7. Powers: $\lim\limits_{x\to c} f(x)^n=L^n$
8. Roots: $\lim\limits_{x\to c} \sqrt[n]{f(x)}=\sqrt[n]{L}$
9. Compositions: Adjust our previously given limit situation to:

$$\lim_{x\to c} f(x)=L,\ \lim_{x\to L} g(x)=K \text{ and } g(L)=K. \tag{1.3.1}$$

$$\text{Then } \lim_{x\to c} g(f(x))=K. \tag{1.3.2}$$

We make a note about Property #8: when n is even, L must be greater than 0. If n is odd, then the statement is true for all L.

We apply the theorem to an example.

Example 1.3.1: Using Basic Limit Properties

Let

$$\lim_{x\to 2} f(x)=2,\quad \lim_{x\to 2} g(x)=3\quad \text{and}\quad p(x)=3x^2-5x+7. \tag{1.3.3}$$

Find the following limits:

1. $\lim\limits_{x\to 2}\big(f(x)+g(x)\big)$
2. $\lim\limits_{x\to 2}\big(5f(x)+g(x)^2\big)$
3. $\lim\limits_{x\to 2} p(x)$

Solution

1. Using the Sum/Difference rule, we know that $\lim\limits_{x\to 2}\big(f(x)+g(x)\big)=2+3=5$.

2. Using the Scalar Multiple and Sum/Difference rules, we find that $\lim\limits_{x\to 2}\left(5f(x)+g(x)^2\right)=5\cdot 2+3^2=19$.
3. Here we combine the Power, Scalar Multiple, Sum/Difference and Constant Rules. We show quite a few steps, but in general these can be omitted:

$$\begin{aligned}\lim_{x\to 2} p(x) &= \lim_{x\to 2}(3x^2-5x+7)\\ &= \lim_{x\to 2} 3x^2 - \lim_{x\to 2} 5x + \lim_{x\to 2} 7\\ &= 3\cdot 2^2 - 5\cdot 2+7\\ &= 9\end{aligned}$$

Part 3 of the previous example demonstrates how the limit of a quadratic polynomial can be determined using the properties of Theorem 1. Not only that, recognize that

$$\lim_{x\to 2} p(x) = 9 = p(2); \tag{1.3.4}$$

i.e., the limit at 2 was found just by plugging 2 into the function. This holds true for all polynomials, and also for rational functions (which are quotients of polynomials), as stated in the following theorem.

Theorem $1.3.2$: Limits of Polynomial and Rational Functions

Limits of Polynomial and Rational Functions}{Let $p(x)$ and $q(x)$ be polynomials and c a real number. Then:

1. $\lim\limits_{x\to c} p(x) = p(c)$
2. $\lim\limits_{x\to c} \frac{p(x)}{q(x)} = \frac{p(c)}{q(c)}$, where $q(c)\neq 0$.

Example $1.3.2$: Finding a limit of a rational function

Using Theorem 2, find

$$\lim_{x\to -1} \frac{3x^2-5x+1}{x^4-x^2+3}.$$

Solution

Using Theorem $1.3.2$, we can quickly state that

$$\begin{aligned}\lim_{x\to -1} \frac{3x^2-5x+1}{x^4-x^2+3} &= \frac{3(-1)^2-5(-1)+1}{(-1)^4-(-1)^2+3}\\ &= \frac{9}{3}=3.\end{aligned}$$

It was likely frustrating in Section 1.2 to do a lot of work to prove that

$$\lim_{x\to 2} x^2 = 4 \tag{1.3.5}$$

as it seemed fairly obvious. The previous theorems state that many functions behave in such an "obvious" fashion, as demonstrated by the rational function in Example 11.

Polynomial and rational functions are not the only functions to behave in such a predictable way. The following theorem gives a list of functions whose behavior is particularly "nice" in terms of limits. In the next section, we will give a formal name to these functions that behave "nicely."

Theorem $1.3.4$: Special Limits

Let c be a real number in the domain of the given function and let n be a positive integer. The following limits hold:

$$1. \lim_{x\to c} \sin x = \sin c \qquad 4. \lim_{x\to c} \csc x = \csc c \qquad 7. \lim_{x\to c} a^x = a^c (a>0) \tag{1.3.6}$$

$$2. \lim_{x\to c} \cos x = \cos c \qquad 5. \lim_{x\to c} \sec x = \sec c \qquad 8. \lim_{x\to c} \ln x = \ln c \tag{1.3.7}$$

$$3. \lim_{x\to c} \tan x = \tan c \qquad 6. \lim_{x\to c} \cot x = \cot c \qquad 9. \lim_{x\to c} \sqrt[n]{x} = \sqrt[n]{c} \tag{1.3.8}$$

Example 1.3.3: Evaluating limits analytically

Evaluate the following limits.

1. $\lim_{x\to\pi}\cos x$
2. $\lim_{x\to 3}(\sec^2 x-\tan^2 x)$
3. $\lim_{x\to\pi/2}\cos x\sin x$
4. $\lim_{x\to 1}e^{\ln x}$
5. $\lim_{x\to 0}\frac{\sin x}{x}$

Solution

1. This is a straightforward application of Theorem 3. $\lim_{x\to\pi}\cos x=\cos\pi=-1$.
2. We can approach this in at least two ways. First, by directly applying Theorem 3, we have:

   $$\lim_{x\to 3}(\sec^2 x-\tan^2 x)=\sec^2 3-\tan^2 3. \tag{1.3.9}$$

   Using the Pythagorean Theorem, this last expression is 1; therefore

   $$\lim_{x\to 3}(\sec^2 x-\tan^2 x)=1. \tag{1.3.10}$$

   We can also use the Pythagorean Theorem from the start.

   $$\lim_{x\to 3}(\sec^2 x-\tan^2 x)=\lim_{x\to 3}1=1, \tag{1.3.11}$$

   using the Constant limit rule. Either way, we find the limit is 1.
3. Applying the Product limit rule of Theorem 1 and Theorem 3 gives

   $$\lim_{x\to\pi/2}\cos x\sin x=\cos(\pi/2)\sin(\pi/2)=0\cdot 1=0. \tag{1.3.12}$$

4. Again, we can approach this in two ways. First, we can use the exponential/logarithmic identity that $e^{\ln x}=x$ and evaluate $\lim_{x\to 1}e^{\ln x}=\lim_{x\to 1}x=1$.
   We can also use the limit Composition Rule of Theorem 1. Using Theorem 3, we have $\lim_{x\to 1}\ln x=\ln 1=0$ and $\lim_{x\to 0}e^x=e^0$ $=1$, satisfying the conditions of the Composition Rule. Applying this rule,

   $$\lim_{x\to 1}e^{\ln x}=\lim_{x\to 0}e^x=e^0=1. \tag{1.3.13}$$

   Both approaches are valid, giving the same result.
5. We encountered this limit in Section 1.1. Applying our theorems, we attempt to find the limit as

   $$\lim_{x\to 0}\frac{\sin x}{x}\to\frac{\sin 0}{0}\to\frac{''0''}{0}. \tag{1.3.14}$$

   This, of course, violates a condition of Theorem 1, as the limit of the denominator is not allowed to be 0. Therefore, we are still unable to evaluate this limit with tools we currently have at hand.

The section could have been titled "Using Known Limits to Find Unknown Limits." By knowing certain limits of functions, we can find limits involving sums, products, powers, etc., of these functions. We further the development of such comparative tools with the Squeeze Theorem, a clever and intuitive way to find the value of some limits.

Before stating this theorem formally, suppose we have functions f, g and h where g always takes on values between f and h; that is, for all x in an interval,

$$f(x)\le g(x)\le h(x). \tag{1.3.15}$$

If f and h have the same limit at c, and g is always "squeezed" between them, then g must have the same limit as well. That is what the Squeeze Theorem states.

Theorem 1.3.5: Squeeze Theorem

Let f, g and h be functions on an open interval I containing c such that for all x in I,

$$f(x) \leq g(x) \leq h(x). \tag{1.3.16}$$

If

$$\lim_{x\to c} f(x) = L = \lim_{x\to c} h(x), \tag{1.3.17}$$

then

$$\lim_{x\to c} g(x) = L. \tag{1.3.18}$$

It can take some work to figure out appropriate functions by which to "squeeze" the given function of which you are trying to evaluate a limit. However, that is generally the only place work is necessary; the theorem makes the "evaluating the limit part" very simple.

We use the Squeeze Theorem in the following example to finally prove that $\lim\limits_{x\to 0} \frac{\sin x}{x} = 1$.

Example 1.3.4: Using the Squeeze Theorem

Use the Squeeze Theorem to show that

$$\lim_{x\to 0} \frac{\sin x}{x} = 1.$$

Solution

We begin by considering the unit circle. Each point on the unit circle has coordinates $(\cos\theta, \sin\theta)$ for some angle θ as shown in Figure $1.3.1$. Using similar triangles, we can extend the line from the origin through the point to the point $(1, \tan\theta)$, as shown. (Here we are assuming that $0 \leq \theta \leq \pi/2$.Later we will show that we can also consider $\theta \leq 0$.)

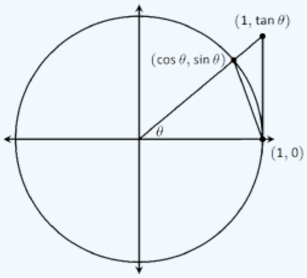
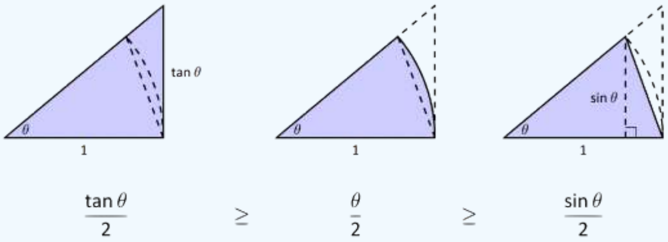


Figure $1.3.1$: The unit circle and related triangles.

Figure 1.19 shows three regions have been constructed in the first quadrant, two triangles and a sector of a circle, which are also drawn below. The area of the large triangle is $\frac{1}{2}\tan\theta$; the area of the sector is $\theta/2$; the area of the triangle contained inside the sector is $\frac{1}{2}\sin\theta$. It is then clear from the diagram that



Multiply all terms by $\frac{2}{\sin\theta}$, giving

$$\frac{1}{\cos\theta} \geq \frac{\theta}{\sin\theta} \geq 1. \tag{1.3.19}$$

Taking reciprocals reverses the inequalities, giving

$$\cos\theta \leq \frac{\sin\theta}{\theta} \leq 1. \tag{1.3.20}$$

(These inequalities hold for all values of θ near 0, even negative values, since $\cos(-\theta)=\cos\theta$ and $\sin(-\theta)=-\sin\theta$.)

Now take limits.

$$\lim_{\theta\to 0}\cos\theta \leq \lim_{\theta\to 0}\frac{\sin\theta}{\theta} \leq \lim_{\theta\to 0} 1 \tag{1.3.21}$$

$$\cos 0 \leq \lim_{\theta\to 0}\frac{\sin\theta}{\theta} \leq 1 \tag{1.3.22}$$

$$1 \leq \lim_{\theta\to 0}\frac{\sin\theta}{\theta} \leq 1 \tag{1.3.23}$$

Clearly this means that $\lim\limits_{\theta\to 0}\frac{\sin\theta}{\theta}=1$

Two notes about the previous example are worth mentioning. First, one might be discouraged by this application, thinking "I would *never* have come up with that on my own. This is too hard!" Don't be discouraged; within this text we will guide you in your use of the Squeeze Theorem. As one gains mathematical maturity, clever proofs like this are easier and easier to create.

Second, this limit tells us more than just that as x approaches 0, $\sin(x)/x$ approaches 1. Both x and $\sin x$ are approaching 0, but the *ratio* of x and $\sin x$ approaches 1, meaning that they are approaching 0 in essentially the same way. Another way of viewing this is: for small x, the functions $y=x$ and $y=\sin x$ are essentially indistinguishable.

We include this special limit, along with three others, in the following theorem.

Theorem $1.3.5$: Special Limits

1. $\lim\limits_{x\to 0}\frac{\sin x}{x}=1$
2. $\lim\limits_{x\to 0}\frac{\cos x-1}{x}=0$
3. $\lim\limits_{x\to 0}(1+x)^{\frac{1}{x}}=e$
4. $\lim\limits_{x\to 0}\frac{e^x-1}{x}=1$

A short word on how to interpret the latter three limits. We know that as x goes to 0, $\cos x$ goes to 1. So, in the second limit, both the numerator and denominator are approaching 0. However, since the limit is 0, we can interpret this as saying that "$\cos x$ is approaching 1 faster than x is approaching 0."

In the third limit, inside the parentheses we have an expression that is approaching 1 (though never equaling 1), and we know that 1 raised to any power is still 1. At the same time, the power is growing toward infinity. What happens to a number near 1 raised to a very large power? In this particular case, the result approaches Euler's number, e, approximately 2.718.

In the fourth limit, we see that as $x\to 0$, e^x approaches 1 ``just as fast" as $x\to 0$, resulting in a limit of 1.

Our final theorem for this section will be motivated by the following example.

Example $1.3.5$: Using algebra to evaluate a limit

Evaluate the following limit:

$$\lim_{x\to 1}\frac{x^2-1}{x-1}.$$

Solution

We begin by attempting to apply Theorem 3 and substituting 1 for x in the quotient. This gives:

$$\lim_{x\to 1}\frac{x^2-1}{x-1}=\frac{1^2-1}{1-1}=\frac{''0''}{0}, \tag{1.3.24}$$

and indeterminate form. We cannot apply the theorem.

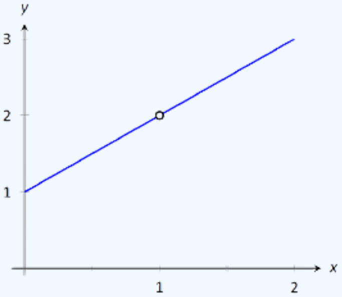


Figure $1.3.2$: Graphing f in Example $1.3.5$ to understand a limit.

By graphing the function, as in Figure $1.3.2$, we see that the function seems to be linear, implying that the limit should be easy to evaluate. Recognize that the numerator of our quotient can be factored:

$$\frac{x^2-1}{x-1}=\frac{(x-1)(x+1)}{x-1}.$$

The function is not defined when $x=1$, but for all other x,

$$\frac{x^2-1}{x-1}=\frac{(x-1)(x+1)}{x-1}=\frac{\cancel{(x-1)}\,(x+1)}{\cancel{x-1}}=x+1. \tag{1.3.25}$$

Clearly $\lim\limits_{x\to 1} x+1=2$. Recall that when considering limits, we are not concerned with the value of the function at 1, only the value the function approaches as x approaches 1. Since $(x^2-1)/(x-1)$ and $x+1$ are the same at all points except $x=1$, they both approach the same value as x approaches 1. Therefore we can conclude that

$$\lim_{x\to 1}\frac{x^2-1}{x-1}=2.$$

The key to the above example is that the functions $y=(x^2-1)/(x-1)$ and $y=x+1$ are identical except at $x=1$. Since limits describe a value the function is approaching, not the value the function actually attains, the limits of the two functions are always equal.

Theorem 1.3.6 : Limits of Functions Equal At All But One Point

Let $g(x)=f(x)$ for all x in an open interval, except possibly at c, and let $\lim\limits_{x\to c} g(x)=L$ for some real number L.Then

$$\lim_{x\to c} f(x)=L. \tag{1.3.26}$$

The Fundamental Theorem of Algebra tells us that when dealing with a rational function of the form $g(x)/f(x)$ and directly evaluating the limit $\lim\limits_{x\to c}\frac{g(x)}{f(x)}$ returns "0/0", then $(x-c)$ is a factor of both $g(x)$ and $f(x)$. One can then use algebra to factor this term out, cancel, then apply Theorem 6. We demonstrate this once more.

Example $1.3.6$: Evaluating a limit using Theorem $1.3.6$

Evaluate

$$\lim_{x\to 3}\frac{x^3-2x^2-5x+6}{2x^3+3x^2-32x+15}.$$

Solution

We begin by applying Theorem 3 and substituting 3 for x.This returns the familiar indeterminate form of "0/0".

Since the numerator and denominator are each polynomials, we know that $(x-3)$ is factor of each. Using whatever method is most comfortable to you, factor out $(x-3)$ from each (using polynomial division, synthetic division, a computer algebra system, etc.). We find that

$$\frac{x^3-2x^2-5x+6}{2x^3+3x^2-32x+15}=\frac{(x-3)(x^2+x-2)}{(x-3)(2x^2+9x-5)}.$$

We can cancel the $(x-3)$ terms as long as $x\neq 3$. Using Theorem 6 we conclude:

$$\begin{aligned}\lim_{x\to 3}\frac{x^3-2x^2-5x+6}{2x^3+3x^2-32x+15} &= \lim_{x\to 3}\frac{(x-3)(x^2+x-2)}{(x-3)(2x^2+9x-5)}\\ &= \lim_{x\to 3}\frac{(x^2+x-2)}{(2x^2+9x-5)}\\ &= \frac{10}{40}=\frac{1}{4}.\end{aligned}$$

We end this section by revisiting a limit first seen in Section 1.1, a limit of a difference quotient. Let $f(x)=-1.5x^2+11.5x$; we approximated the limit $\lim_{h\to 0}\frac{f(1+h)-f(1)}{h}\approx 8.5$. We formally evaluate this limit in the following example.

Example $1.3.7$: Evaluating the limit of a difference quotient

Let $f(x)=-1.5x^2+11.5x$; find $\lim_{h\to 0}\frac{f(1+h)-f(1)}{h}$.

Solution

Since f is a polynomial, our first attempt should be to employ Theorem 3 and substitute 0 for h. However, we see that this gives us "0/0." Knowing that we have a rational function hints that some algebra will help. Consider the following steps:

$$\begin{aligned}\lim_{h\to 0}\frac{f(1+h)-f(1)}{h} &= \lim_{h\to 0}\frac{-1.5(1+h)^2+11.5(1+h)-(-1.5(1)^2+11.5(1))}{h}\\ &= \lim_{h\to 0}\frac{-1.5(1+2h+h^2)+11.5+11.5h-10}{h}\\ &= \lim_{h\to 0}\frac{-1.5h^2+8.5h}{h}\\ &= \lim_{h\to 0}\frac{h(-1.5h+8.5)}{h}\\ &= \lim_{h\to 0}(-1.5h+8.5) \quad (\text{using Theorem 6, as } h\neq 0)\\ &= 8.5 \quad (\text{using Theorem 3})\end{aligned}$$

This matches our previous approximation.

This section contains several valuable tools for evaluating limits. One of the main results of this section is Theorem 3; it states that many functions that we use regularly behave in a very nice, predictable way. In the next section we give a name to this nice behavior; we label such functions as *continuous*. Defining that term will require us to look again at what a limit is and what causes limits to not exist.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 1.3: Finding Limits Analytically is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

1.4: One Sided Limits

We introduced the concept of a limit gently, approximating their values graphically and numerically. Next came the rigorous definition of the limit, along with an admittedly tedious method for evaluating them. The previous section gave us tools (which we call theorems) that allow us to compute limits with greater ease. Chief among the results were the facts that polynomials and rational, trigonometric, exponential and logarithmic functions (and their sums, products, etc.) all behave "nicely." In this section we rigorously define what we mean by "nicely."

In Section 1.1 we explored the three ways in which limits of functions failed to exist:

1. The function approached different values from the left and right,
2. The function grows without bound, and
3. The function oscillates.

In this section we explore in depth the concepts behind #1 by introducing the *one-sided limit*. We begin with formal definitions that are very similar to the definition of the limit given in Section 1.2, but the notation is slightly different and "$x \neq c$" is replaced with either "$x < c$" or "$x > c$."

Definition 2: One Sided Limits

Left-Hand Limit

Let I be an open interval containing c, and let f be a function defined on I, except possibly at c. The **limit of** $f(x)$, **as** x **approaches** c **from the left**, is L, or, **the left--hand limit of** f **at** c is L, denoted by

$$\lim_{x \to c^-} f(x) = L, \tag{1.4.1}$$

means that given any $\epsilon > 0$, there exists $\delta > 0$ such that for all $x < c$, if $|x - c| < \delta$, then $|f(x) - L| < \epsilon$.

Right-Hand Limit

Let I be an open interval containing c, and let f be a function defined on I, except possibly at c. The **limit of** $f(x)$, **as** x **approaches** c **from the right, is** L, or, **the right--hand limit of** f **at** c **is** L, denoted by

$$\lim_{x \to c^+} f(x) = L, \tag{1.4.2}$$

means that given any $\epsilon > 0$, there exists $\delta > 0$ such that for all $x > c$, if $|x - c| < \delta$, then $|f(x) - L| < \epsilon$.

Practically speaking, when evaluating a left-hand limit, we consider only values of x "to the left of c," i.e., where $x < c$. The admittedly imperfect notation $x \to c^-$ is used to imply that we look at values of x to the left of c. The notation has nothing to do with positive or negative values of either x or c. A similar statement holds for evaluating right-hand limits; there we consider only values of x to the right of c, i.e., $x > c$. We can use the theorems from previous sections to help us evaluate these limits; we just restrict our view to one side of c.

We practice evaluating left and right-hand limits through a series of examples.

Example 17: Evaluating one sided limits

Let $f(x) = \begin{cases} x & 0 \leq x \leq 1 \\ 3 - x & 1 < x < 2 \end{cases}$, as shown in Figure 1.21. Find each of the following:

1. $\lim\limits_{x \to 1^-} f(x)$
2. $\lim\limits_{x \to 1^+} f(x)$
3. $\lim\limits_{x \to 1} f(x)$
4. $f(1)$
5. $\lim\limits_{x \to 0^+} f(x)$
6. $f(0)$
7. $\lim\limits_{x \to 2^-} f(x)$
8. $f(2)$

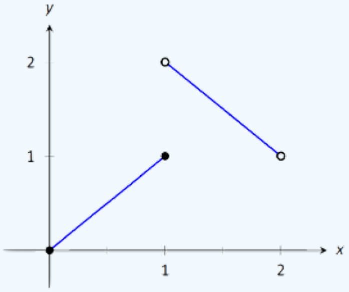


FIGURE 1.21: A graph of f in Example 17.

Solution

For these problems, the visual aid of the graph is likely more effective in evaluating the limits than using f itself. Therefore we will refer often to the graph.

1. As x goes to 1 *from the left*, we see that $f(x)$ is approaching the value of 1. Therefore $\lim_{x\to 1^-} f(x) = 1$.
2. As x goes to 1 *from the right*, we see that $f(x)$ is approaching the value of 2. Recall that it does not matter that there is an "open circle" there; we are evaluating a limit, not the value of the function. Therefore $\lim_{x\to 1^+} f(x) = 2$.
3. *The* limit of f as x approaches 1 does not exist, as discussed in the first section. The function does not approach one particular value, but two different values from the left and the right.
4. Using the definition and by looking at the graph we see that $f(1) = 1$.
5. As x goes to 0 from the right, we see that $f(x)$ is also approaching 0. Therefore $\lim_{x\to 0^+} f(x) = 0$. Note we cannot consider a left-hand limit at 0 as f is not defined for values of $x < 0$.
6. Using the definition and the graph, $f(0) = 0$.
7. As x goes to 2 from the left, we see that $f(x)$ is approaching the value of 1. Therefore $\lim_{x\to 2^-} f(x) = 1$.
8. The graph and the definition of the function show that $f(2)$ is not defined.

Note how the left and right-hand limits were different at $x = 1$. This, of course, causes *the* limit to not exist. The following theorem states what is fairly intuitive: *the* limit exists precisely when the left and right-hand limits are equal.

Theorem 7: Limits and One Sided Limits

Let f be a function defined on an open interval I containing c. Then

$$\lim_{x\to c} f(x) = L \tag{1.4.3}$$

if, and only if,

$$\lim_{x\to c^-} f(x) = L \quad \text{and} \quad \lim_{x\to c^+} f(x) = L. \tag{1.4.4}$$

The phrase "if, and only if" means the two statements are *equivalent*: they are either both true or both false. If the limit equals L, then the left and right hand limits both equal L. If the limit is not equal to L, then at least one of the left and right-hand limits is not equal to L (it may not even exist).

One thing to consider in Examples 17 - 20 is that the value of the function may/may not be equal to the value(s) of its left/right-hand limits, even when these limits agree.

Example 18: Evaluating limits of a piecewise-defined function

Let $f(x) = \begin{cases} 2-x & 0<x<1 \\ (x-2)^2 & 1<x<2 \end{cases}$, as shown in Figure 1.22. Evaluate the following.

1. $\lim_{x\to 1^-} f(x)$
2. $\lim_{x\to 1^+} f(x)$
3. $\lim_{x\to 1} f(x)$

4. $f(1)$
5. $\lim_{x\to 0^+} f(x)$
6. $f(0)$
7. $\lim_{x\to 2^-} f(x)$
8. $f(2)$

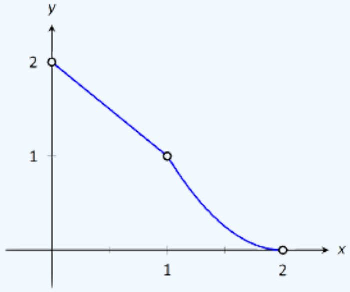


FIGURE 1.22: A graph of f from Example 18.

Solution

Again we will evaluate each using both the definition of f and its graph.

1. As x approaches 1 from the left, we see that $f(x)$ approaches 1. Therefore $\lim_{x\to 1^-} f(x) = 1$.
2. As x approaches 1 from the right, we see that again $f(x)$ approaches 1. Therefore $\lim_{x\to 1+} f(x) = 1$.
3. *The* limit of f as x approaches 1 exists and is 1, as f approaches 1 from both the right and left. Therefore $\lim_{x\to 1} f(x) = 1$.
4. $f(1)$ is not defined. Note that 1 is not in the domain of f as defined by the problem, which is indicated on the graph by an open circle when $x = 1$.
5. As x goes to 0 from the right, $f(x)$ approaches 2. So $\lim_{x\to 0^+} f(x) = 2$.
6. $f(0)$ is not defined as 0 is not in the domain of f.
7. As x goes to 2 from the left, $f(x)$ approaches 0. So $\lim_{x\to 2^-} f(x) = 0$.
8. $f(2)$ is not defined as 2 is not in the domain of f.

Example 19: Evaluating limits of a piecewise-defined function

Let $f(x) = \begin{cases} (x-1)^2 & 0 \le x \le 2, x \ne 1 \\ 1 & x = 1 \end{cases}$, as shown in Figure 1.23. Evaluate the following.

1. $\lim_{x\to 1^-} f(x)$
2. $\lim_{x\to 1^+} f(x)$
3. $\lim_{x\to 1} f(x)$
4. $f(1)$

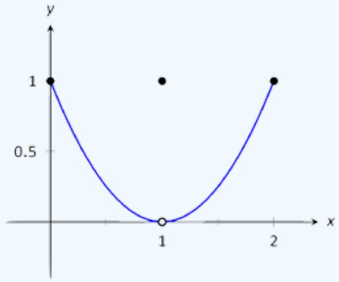


FIGURE 1.23: Graphing f in Example 19.

It is clear by looking at the graph that both the left and right-hand limits of f, as x approaches 1, is 0. Thus it is also clear that *the* limit is 0; i.e., $\lim_{x\to 1} f(x) = 0$. It is also clearly stated that $f(1) = 1$.

Example 20: Evaluating limits of a piecewise-defined function

Let $f(x)=\left\{\begin{array}{cc}x^2 & 0\leq x\leq 1\\ 2-x & 1<x\leq 2\end{array}\right.$, as shown in Figure 1.24. Evaluate the following.

1. $\lim\limits_{x\to 1^-} f(x)$
2. $\lim\limits_{x\to 1^+} f(x)$
3. $\lim\limits_{x\to 1} f(x)$
4. $f(1)$

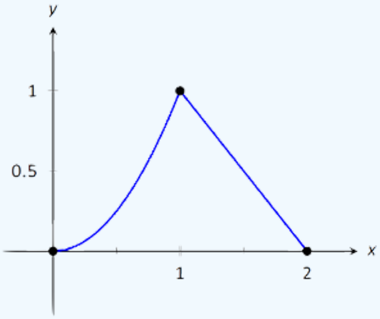


FIGURE 1.24: Graphing f in Example 20.

Solution

It is clear from the definition of the function and its graph that all of the following are equal:

$$\lim_{x\to 1^-} f(x)=\lim_{x\to 1^+} f(x)=\lim_{x\to 1} f(x)=f(1)=1. \tag{1.4.5}$$

In Examples 17 - 20 we were asked to find both $\lim\limits_{x\to 1} f(x)$ and $f(1)$. Consider the following table:

	$\lim\limits_{x\to 1} f(x)$	$f(1)$
Example 17	does not exist	1
Example 18	1	not defined
Example 19	0	1
Example 20	1	1

(1.4.6)

Only in Example 20 do both the function and the limit exist and agree. This seems "nice;" in fact, it seems "normal." This is in fact an important situation which we explore in the next section, entitled "Continuity." In short, a *continuous function* is one in which when a function approaches a value as $x\to c$ (i.e., when $\lim\limits_{x\to c} f(x)=L$), it actually *attains* that value at c. Such functions behave nicely as they are very predictable.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 1.4: One Sided Limits is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

1.5: Continuity

As we have studied limits, we have gained the intuition that limits measure ``where a function is heading." That is, if

$$\lim_{x \to 1} f(x) = 3 \tag{1.5.1}$$

then as x is close to 1, $f(x)$ is close to 3. We have seen, though, that this is not necessarily a good indicator of what $f(1)$ actually is. This can be problematic; functions can tend to one value but attain another. This section focuses on functions that *do not* exhibit such behavior.

Definition 3 Continuous Function

Let f be a function defined on an open interval I containing c.

1. f is **continuous at** c if $\lim_{x \to c} f(x) = f(c)$.
2. f is **continuous on** I if f is continuous at c for all values of c in I. If f is continuous on $(-\infty, \infty)$, we say f is **continuous everywhere**.

A useful way to establish whether or not a function f is continuous at c is to verify the following three things:

1. $\lim_{x \to c} f(x)$ exists,
2. $f(c)$ is defined, and
3. $\lim_{x \to c} f(x) = f(c)$.

Example 21: Finding intervals of continuity

Let f be defined as shown in Figure 1.25. Give the interval(s) on which f is continuous.

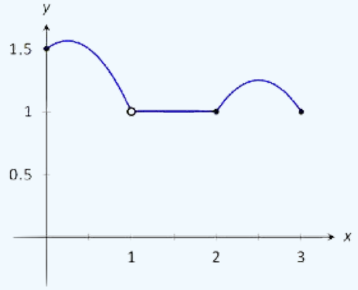


FIGURE 1.25: A graph of f in Example 21.

Solution

We proceed by examining the three criteria for continuity.

1. The limits $\lim_{x \to c} f(x)$ exists for all c between 0 and 3.
2. $f(c)$ is defined for all c between 0 and 3, *except for* $c = 1$. We know immediately that f cannot be continuous at $x = 1$.
3. The limit $\lim_{x \to c} f(x) = f(c)$ for all c between 0 and 3, except, of course, for $c = 1$.

We conclude that f is continuous at every point of $(0, 3)$ except at $x = 1$. Therefore f is continuous on $(0, 1) \cup (1, 3)$.

Example 22: Finding intervals of continuity

The *floor function*, $f(x) = \lfloor x \rfloor$, returns the largest integer smaller than the input x. (For example, $f(\pi) = \lfloor \pi \rfloor = 3$.) The graph of f in Figure 1.26 demonstrates why this is often called a "step function."

Give the intervals on which f is continuous.

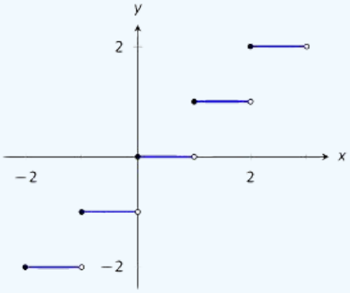


FIGURE 1.26: A graph of the step function in Example 22.

Solution

We examine the three criteria for continuity.

1. The limits $\lim_{x\to c} f(x)$ do not exist at the jumps from one "step" to the next, which occur at all integer values of c. Therefore the limits exist for all c except when c is an integer.
2. The function is defined for all values of c.
3. The limit $\lim_{x\to c} f(x) = f(c)$ for all values of c where the limit exist, since each step consists of just a line.

We conclude that f is continuous everywhere except at integer values of c. So the intervals on which f is continuous are

$$\ldots, (-2,-1), (-1,0), (0,1), (1,2), \ldots. \tag{1.5.2}$$

Our definition of continuity on an interval specifies the interval is an open interval. We can extend the definition of continuity to closed intervals by considering the appropriate one-sided limits at the endpoints.

Definition 4: Continuity on Closed Intervals

Let f be defined on the closed interval $[a,b]$ for some real numbers a, b. f is **continuous on** $[a,b]$ if:

1. f is continuous on (a,b),
2. $\lim_{x\to a^+} f(x) = f(a)$ and
3. $\lim_{x\to b^-} f(x) = f(b)$.

We can make the appropriate adjustments to talk about continuity on half--open intervals such as $[a,b)$ or $(a,b]$ if necessary.

Example 23: Determining intervals on which a function is continuous

For each of the following functions, give the domain of the function and the interval(s) on which it is continuous.

1. $f(x) = 1/x$
2. $f(x) = \sin x$
3. $f(x) = \sqrt{x}$
4. $f(x) = \sqrt{1-x^2}$
5. $f(x) = |x|$

Solution

We examine each in turn.

1. The domain of $f(x) = 1/x$ is $(-\infty,0)\cup(0,\infty)$. As it is a rational function, we apply Theorem 2 to recognize that f is continuous on all of its domain.
2. The domain of $f(x) = \sin x$ is all real numbers, or $(-\infty,\infty)$. Applying Theorem 3 shows that $\sin x$ is continuous everywhere.
3. The domain of $f(x) = \sqrt{x}$ is $[0,\infty)$. Applying Theorem 3 shows that $f(x) = \sqrt{x}$ is continuous on its domain of $[0,\infty)$.
4. The domain of $f(x) = \sqrt{1-x^2}$ is $[-1,1]$. Applying Theorems 1 and 3 shows that f is continuous on all of its domain, $[-1,1]$.

5. The domain of $f(x)=|x|$ is $(-\infty,\infty)$. We can define the absolute value function as $f(x)=\begin{cases}-x & x<0\\ x & x\geq 0\end{cases}$. Each "piece" of this piecewise defined function is continuous on all of its domain, giving that f is continuous on $(-\infty,0)$ and $[0,\infty)$. We cannot assume this implies that f is continuous on $(-\infty,\infty)$; we need to check that $\lim\limits_{x\to 0} f(x)=f(0)$, as $x=0$ is the point where f transitions from one ``piece" of its definition to the other. It is easy to verify that this is indeed true, hence we conclude that $f(x)=|x|$ is continuous everywhere.

Continuity is inherently tied to the properties of limits. Because of this, the properties of limits found in Theorems 1 and 2 apply to continuity as well. Further, now knowing the definition of continuity we can re--read Theorem 3 as giving a list of functions that are continuous on their domains. The following theorem states how continuous functions can be combined to form other continuous functions, followed by a theorem which formally lists functions that we know are continuous on their domains.

Theorem 8: Properties of Continuous Functions

Let f and g be continuous functions on an interval I, let c be a real number and let n be a positive integer. The following functions are continuous on I.

1. Sums/Differences : $f\pm g$
2. Constant Multiples : $c\cdot f$
3. Products : $f\cdot g$
4. Quotients : f/g (as long as $g\neq 0$ on I)
5. Powers : f^n
6. Roots : $\sqrt[n]{f}$ (if n is even then $f\geq 0$ on I; if n is odd, then true for all values of f on I.)
7. Compositions : Adjust the definitions of f and g to: Let f be continuous on I, where the range of f on I is J, and let g be continuous on J. Then $g\circ f$, i.e., $g(f(x))$, is continuous on I.

Theorem 9: Continuous Functions

The following functions are continuous on their domains.

$$1.\ f(x)=\sin x \qquad 2.\ f(x)=\cos x \tag{1.5.3}$$
$$3.\ f(x)=\tan x \qquad 4.\ f(x)=\cot x \tag{1.5.4}$$
$$5.\ f(x)=\sec x \qquad 6.\ f(x)=\csc x \tag{1.5.5}$$
$$7.\ f(x)=\ln x \qquad 8.\ f(x)=\sqrt[n]{x}, \tag{1.5.6}$$
$$9.\ f(x)=a^x (a>0) \qquad \text{(where n is a positive integer)} \tag{1.5.7}$$

We apply these theorems in the following Example.

Example 24: Determining intervals on which a function is continuous

State the interval(s) on which each of the following functions is continuous.

$$1.\ \ f(x)=\sqrt{x-1}+\sqrt{5-x} \qquad 3.\ \ f(x)=\tan x \tag{1.5.8}$$
$$2.\ \ f(x)=x\sin x \qquad 4.\ \ f(x)=\sqrt{\ln x} \tag{1.5.9}$$

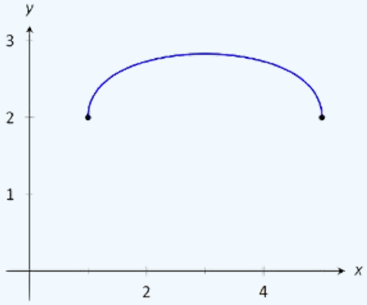


FIGURE 1.27: A graph of f in Example 24(a).

Solution

We examine each in turn, applying Theorems 8 and 9 as appropriate.

1. The square--root terms are continuous on the intervals $[1,\infty)$ and $(-\infty,5]$, respectively. As f is continuous only where each term is continuous, f is continuous on $[1,5]$, the intersection of these two intervals. A graph of f is given in Figure 1.27.
2. The functions $y=x$ and $y=\sin x$ are each continuous everywhere, hence their product is, too.
3. Theorem 9 states that $f(x)=\tan x$ is continuous "on its domain." Its domain includes all real numbers except odd multiples of $\pi/2$. Thus $f(x)=\tan x$ is continuous on

$$\ldots\left(-\frac{3\pi}{2},-\frac{\pi}{2}\right),\ \left(-\frac{\pi}{2},\frac{\pi}{2}\right),\ \left(\frac{\pi}{2},\frac{3\pi}{2}\right),\ldots, \tag{1.5.10}$$

   or, equivalently, on $D=\{x\in\mathbb{R}\mid x\neq n\cdot\frac{\pi}{2},\ \text{n is an odd integer}\}$.
4. The domain of $y=\sqrt{x}$ is $[0,\infty)$. The range of $y=\ln x$ is $(-\infty,\infty)$, but if we restrict its domain to $[1,\infty)$ its range is $[0,\infty)$. So restricting $y=\ln x$ to the domain of $[1,\infty)$ restricts its output is $[0,\infty)$, on which $y=\sqrt{x}$ is defined. Thus the domain of $f(x)=\sqrt{\ln x}$ is $[1,\infty)$.

A common way of thinking of a continuous function is that "its graph can be sketched without lifting your pencil." That is, its graph forms a "continuous" curve, without holes, breaks or jumps. While beyond the scope of this text, this pseudo--definition glosses over some of the finer points of continuity. Very strange functions are continuous that one would be hard pressed to actually sketch by hand.

This intuitive notion of continuity does help us understand another important concept as follows. Suppose f is defined on $[1,2]$ and $f(1)=-10$ and $f(2)=5$. If f is continuous on $[1,2]$ (i.e., its graph can be sketched as a continuous curve from $(1,-10)$ to $(2,5)$) then we know intuitively that somewhere on $[1,2]$ f must be equal to -9, and -8, and $-7,\ -6,\ \ldots,\ 0,\ 1/2$, etc. In short, f takes on all *intermediate* values between -10 and 5. It may take on more values; f may actually equal 6 at some time, for instance, but we are guaranteed all values between -10 and 5.

While this notion seems intuitive, it is not trivial to prove and its importance is profound. Therefore the concept is stated in the form of a theorem.

Theorem 10: Intermediate Value Theorem

Let f be a continuous function on $[a,b]$ and, without loss of generality, let $f(a)<f(b)$. Then for every value y, where $f(a)<y<f(b)$, there is a value c in $[a,b]$ such that $f(c)=y$.

One important application of the Intermediate Value Theorem is root finding. Given a function f, we are often interested in finding values of x where $f(x)=0$. These roots may be very difficult to find exactly. Good approximations can be found through successive applications of this theorem. Suppose through direct computation we find that $f(a)<0$\(and\($f(b)>0$, where $a<b$. The Intermediate Value Theorem states that there is a c in $[a,b]$ such that $f(c)=0$. The theorem does not give us any clue as to where that value is in the interval $[a,b]$, just that it exists.

There is a technique that produces a good approximation of c. Let d be the midpoint of the interval $[a,b]$ and consider $f(d)$. There are three possibilities:

1. $f(d)=0$ -- we got lucky and stumbled on the actual value. We stop as we found a root.
2. $f(d)<0$ Then we know there is a root of f on the interval $[d,b]$ -- we have halved the size of our interval, hence are closer to a good approximation of the root.
3. $f(d)>0$ Then we know there is a root of f on the interval $[a,d]$ -- again,we have halved the size of our interval, hence are closer to a good approximation of the root.

Successively applying this technique is called the **Bisection Method** of root finding. We continue until the interval is sufficiently small. We demonstrate this in the following example.

Example 25: Using the Bisection Method

Approximate the root of $f(x)=x-\cos x$, accurate to three places after the decimal.

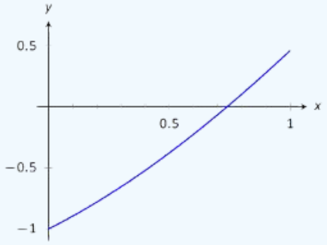
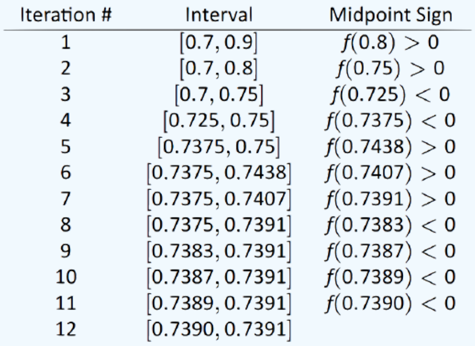


FIGURE 1.28: Graphing a root of $f(x) = x - \cos x$.

Solution

Consider the graph of $f(x) = x - \cos x$, shown in Figure 1.28. It is clear that the graph crosses the x-axis somewhere near $x = 0.8$. To start the Bisection Method, pick an interval that contains 0.8. We choose $[0.7, 0.9]$. Note that all we care about are signs of $f(x)$, not their actual value, so this is all we display.

1. **Iteration 1**: $f(0.7) < 0$, $f(0.9) > 0$, and $f(0.8) > 0$. So replace 0.9 with 0.8 and repeat.
2. **Iteration 2**: $f(0.7) < 0$, $f(0.8) > 0$, and at the midpoint, 0.75, we have $f(0.75) > 0$. So replace 0.8 with 0.75 and repeat. Note that we don't need to continue to check the endpoints, just the midpoint. Thus we put the rest of the iterations in Table 1.29.



Iteration #	Interval	Midpoint Sign
1	$[0.7, 0.9]$	$f(0.8) > 0$
2	$[0.7, 0.8]$	$f(0.75) > 0$
3	$[0.7, 0.75]$	$f(0.725) < 0$
4	$[0.725, 0.75]$	$f(0.7375) < 0$
5	$[0.7375, 0.75]$	$f(0.7438) > 0$
6	$[0.7375, 0.7438]$	$f(0.7407) > 0$
7	$[0.7375, 0.7407]$	$f(0.7391) > 0$
8	$[0.7375, 0.7391]$	$f(0.7383) < 0$
9	$[0.7383, 0.7391]$	$f(0.7387) < 0$
10	$[0.7387, 0.7391]$	$f(0.7389) < 0$
11	$[0.7389, 0.7391]$	$f(0.7390) < 0$
12	$[0.7390, 0.7391]$	

FIGURE 1.29: Iterations of the Bisection Method of Root Finding

Notice that in the 12$^{\text{th}}$ iteration we have the endpoints of the interval each starting with 0.739. Thus we have narrowed the zero down to an accuracy of the first three places after the decimal. Using a computer, we have

$$f(0.7390) = -0.00014, \quad f(0.7391) = 0.000024. \tag{1.5.11}$$

Either endpoint of the interval gives a good approximation of where f is 0. The Intermediate Value Theorem states that the actual zero is still within this interval. While we do not know its exact value, we know it starts with 0.739.

This type of exercise is rarely done by hand. Rather, it is simple to program a computer to run such an algorithm and stop when the endpoints differ by a preset small amount. One of the authors did write such a program and found the zero of f, accurate to 10 places after the decimal, to be 0.7390851332. While it took a few minutes to write the program, it took less than a thousandth of a second for the program to run the necessary 35 iterations. In less than 8 hundredths of a second, the zero was calculated to 100 decimal places (with less than 200 iterations).

It is a simple matter to extend the Bisection Method to solve problems similar to "Find x, where $f(x) = 0$." For instance, we can find x, where $f(x) = 1$. It actually works very well to define a new function g where $g(x) = f(x) - 1$. Then use the Bisection Method to solve $g(x) = 0$.

Similarly, given two functions f and g, we can use the Bisection Method to solve $f(x) = g(x)$. Once again, create a new function h where $h(x) = f(x) - g(x)$ and solve $h(x) = 0$.

In Section 4.1 another equation solving method will be introduced, called Newton's Method. In many cases, Newton's Method is much faster. It relies on more advanced mathematics, though, so we will wait before introducing it.

This section formally defined what it means to be a continuous function. "Most" functions that we deal with are continuous, so often it feels odd to have to formally define this concept. Regardless, it is important, and forms the basis of the next chapter.

In the next section we examine one more aspect of limits: limits that involve infinity.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 1.5: Continuity is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

1.6: Limits Involving Infinity

In Definition 1 we stated that in the equation $\lim\limits_{x\to c} f(x) = L$, both c and L were numbers. In this section we relax that definition a bit by considering situations when it makes sense to let c and/or L be "infinity."

As a motivating example, consider $f(x) = 1/x^2$, as shown in Figure 1.30. Note how, as x approaches 0, $f(x)$ grows very, very large. It seems appropriate, and descriptive, to state that

$$\lim_{x\to 0} \frac{1}{x^2} = \infty. \tag{1.6.1}$$

Also note that as x gets very large, $f(x)$ gets very, very small. We could represent this concept with notation such as

$$\lim_{x\to \infty} \frac{1}{x^2} = 0. \tag{1.6.2}$$

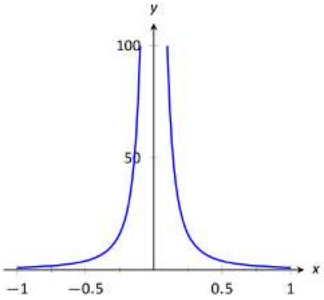


FIGURE 1.30: Graphing $f(x) = 1/x^2$ for values of x near 0.

We explore both types of use of ∞ in turn.

Definition 5: Limit of infinity

We say $\lim\limits_{x\to c} f(x) = \infty$ if for every $M > 0$ there exists $\delta > 0$ such that for all $x \neq c$, if $|x - c| < \delta$, then $f(x) \geq M$.

This is just like the ϵ--δ definition from Section 1.2. In that definition, given any (small) value ϵ, if we let x get close enough to c (within δ units of c) then $f(x)$ is guaranteed to be within ϵ of $f(c)$. Here, given any (large) value M, if we let x get close enough to c (within δ units of c), then $f(x)$ will be at least as large as M. In other words, if we get close enough to c, then we can make $f(x)$ as large as we want. We can define limits equal to $-\infty$ in a similar way.

It is important to note that by saying $\lim\limits_{x\to c} f(x) = \infty$ we are implicitly stating that \textit{the} limit of $f(x)$, as x approaches c, *does not exist*. A limit only exists when $f(x)$ approaches an actual numeric value. We use the concept of limits that approach infinity because it is helpful and descriptive.

Example 26: Evaluating limits involving infinity

Find $\lim\limits_{x\to 1} \frac{1}{(x-1)^2}$ as shown in Figure 1.31.

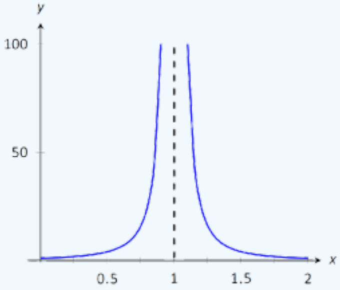


FIGURE 1.31: Observing infinite limit as $x \to 1$ in Example 26.

Solution

In Example 4 of Section 1.1, by inspecting values of x close to 1 we concluded that this limit does not exist. That is, it cannot equal any real number. But the limit could be infinite. And in fact, we see that the function does appear to be growing larger and larger, as $f(.99) = 10^4$, $f(.999) = 10^6$, $f(.9999) = 10^8$. A similar thing happens on the other side of 1. In general, let a "large" value M be given. Let $\delta = 1/\sqrt{M}$. If x is within δ of 1, i.e., if $|x-1| < 1/\sqrt{M}$, then:

$$\begin{aligned} |x-1| &< \frac{1}{\sqrt{M}} \\ (x-1)^2 &< \frac{1}{M} \\ \frac{1}{(x-1)^2} &> M, \end{aligned}$$

which is what we wanted to show. So we may say $\lim\limits_{x\to1} 1/(x-1)^2 = \infty$.

Example 27: Evaluating limits involving infinity

Find $\lim\limits_{x\to0} \frac{1}{x}$, as shown in Figure 1.32.

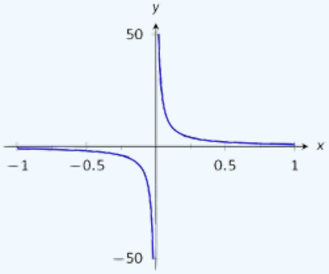


FIGURE 1.32: Evaluating $\lim\limits_{x\to0} \frac{1}{x}$.

Solution

It is easy to see that the function grows without bound near 0, but it does so in different ways on different sides of 0. Since its behavior is not consistent, we cannot say that $\lim\limits_{x\to0} \frac{1}{x} = \infty$. However, we can make a statement about one--sided limits. We can state that $\lim\limits_{x\to0^+} \frac{1}{x} = \infty$ and $\lim\limits_{x\to0^-} \frac{1}{x} = -\infty$.

Vertical Asymptotes

If the limit of $f(x)$ as x approaches c from either the left or right (or both) is ∞ or $-\infty$, we say the function has a **vertical asymptote** at c.

Example 28: Finding vertical asymptotes

Find the vertical asymptotes of $f(x) = \dfrac{3x}{x^2-4}$.

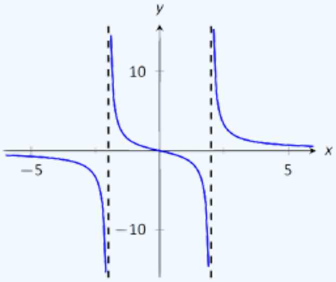


FIGURE 1.33: Graphing $f(x) = \frac{3x}{x^2-4}$.

Solution

> Vertical asymptotes occur where the function grows without bound; this can occur at values of c where the denominator is 0. When x is near c, the denominator is small, which in turn can make the function take on large values. In the case of the given function, the denominator is 0 at $x = \pm 2$. Substituting in values of x close to 2 and -2 seems to indicate that the function tends toward ∞ or $-\infty$ at those points. We can graphically confirm this by looking at Figure 1.33. Thus the vertical asymptotes are at $x = \pm 2$.

When a rational function has a vertical asymptote at $x = c$, we can conclude that the denominator is 0 at $x = c$. However, just because the denominator is 0 at a certain point does not mean there is a vertical asymptote there. For instance, $f(x) = (x^2 - 1)/(x - 1)$ does not have a vertical asymptote at $x = 1$, as shown in Figure 1.34. While the denominator does get small near $x = 1$, the numerator gets small too, matching the denominator step for step. In fact, factoring the numerator, we get

$$f(x) = \frac{(x-1)(x+1)}{x-1}. \tag{1.6.3}$$

Canceling the common term, we get that $f(x) = x + 1$ for $x \neq 1$. So there is clearly no asymptote, rather a hole exists in the graph at $x = 1$.

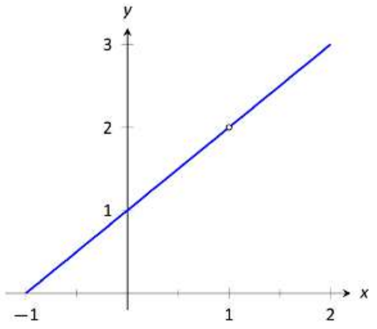


FIGURE 1.34: Graphically showing that $f(x) = \frac{x^2-1}{x-1}$ does not have an asymptote at $x = 1$.

The above example may seem a little contrived. Another example demonstrating this important concept is $f(x) = (\sin x)/x$. We have considered this function several times in the previous sections. We found that $\lim\limits_{x \to 0} \frac{\sin x}{x} = 1$; i.e., there is no vertical asymptote. No simple algebraic cancellation makes this fact obvious; we used the Squeeze Theorem in Section 1.3 to prove this.

If the denominator is 0 at a certain point but the numerator is not, then there will usually be a vertical asymptote at that point. On the other hand, if the numerator and denominator are both zero at that point, then there may or may not be a vertical asymptote at that point. This case where the numerator and denominator are both zero returns us to an important topic.

Indeterminate Forms

We have seen how the limits

$$\lim_{x \to 0} \frac{\sin x}{x} \quad \text{and} \quad \lim_{x \to 1} \frac{x^2 - 1}{x - 1} \tag{1.6.4}$$

each return the indeterminate form "$0/0$" when we blindly plug in $x = 0$ and $x = 1$, respectively. However, $0/0$ is not a valid arithmetical expression. It gives no indication that the respective limits are 1 and 2.

With a little cleverness, one can come up $0/0$ expressions which have a limit of ∞, 0, or any other real number. That is why this expression is called *indeterminate*.

A key concept to understand is that such limits do not really return $0/0$. Rather, keep in mind that we are taking *limits*. What is really happening is that the numerator is shrinking to 0 while the denominator is also shrinking to 0. The respective rates at which they do this are very important and determine the actual value of the limit.

An indeterminate form indicates that one needs to do more work in order to compute the limit. That work may be algebraic (such as factoring and canceling) or it may require a tool such as the Squeeze Theorem. In a later section we will learn a technique called l'Hospital's Rule that provides another way to handle indeterminate forms.

Some other common indeterminate forms are $\infty - \infty$, $\infty \cdot 0$, ∞/∞, 0^0, ∞^0 and 1^∞. Again, keep in mind that these are the "blind" results of evaluating a limit, and each, in and of itself, has no meaning. The expression $\infty - \infty$ does not really mean "subtract infinity from infinity." Rather, it means "One quantity is subtracted from the other, but both are growing without bound." What is the result? It is possible to get every value between $-\infty$ and ∞

Note that $1/0$ and $\infty/0$ are not indeterminate forms, though they are not exactly valid mathematical expressions, either. In each, the function is growing without bound, indicating that the limit will be ∞, $-\infty$, or simply not exist if the left- and right-hand limits do not match.

Limits at Infinity and Horizontal Asymptotes

At the beginning of this section we briefly considered what happens to $f(x) = 1/x^2$ as x grew very large. Graphically, it concerns the behavior of the function to the "far right" of the graph. We make this notion more explicit in the following definition.

Definition 6: Limits at Infinity and Horizontal Asymptote

1. We say $\lim\limits_{x\to\infty} f(x) = L$ if for every $\epsilon > 0$ there exists $M > 0$ such that if $x \geq M$, then $|f(x) - L| < \epsilon$.
2. We say $\lim\limits_{x\to-\infty} f(x) = L$ if for every $\epsilon > 0$ there exists $M < 0$ such that if $x \leq M$, then $|f(x) - L| < \epsilon$.
3. If $\lim\limits_{x\to\infty} f(x) = L$ or $\lim\limits_{x\to-\infty} f(x) = L$, we say that $y = L$ is a **horizontal asymptote** of f.

We can also define limits such as $\lim\limits_{x\to\infty} f(x) = \infty$ by combining this definition with Definition 5.

Example 29: Approximating horizontal asymptotes

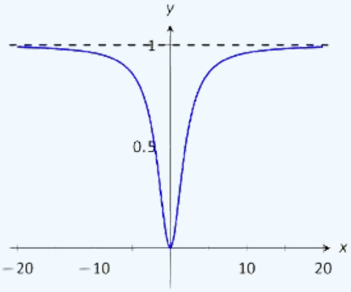
Approximate the horizontal asymptote(s) of $f(x) = \frac{x^2}{x^2+4}$.

Solution

We will approximate the horizontal asymptotes by approximating the limits

$$\lim_{x\to-\infty} \frac{x^2}{x^2+4} \quad \text{and} \quad \lim_{x\to\infty} \frac{x^2}{x^2+4}. \tag{1.6.5}$$

Figure 1.35(a) shows a sketch of f, and part (b) gives values of $f(x)$ for large magnitude values of x. It seems reasonable to conclude from both of these sources that f has a horizontal asymptote at $y = 1$.



(a)

x	$f(x)$
10	0.9615
100	0.9996
10000	0.999996
−10	0.9615
−100	0.9996
−10000	0.999996

(b)

FIGURE 1.35: Using a graph and a table to approximate a horizontal asymptote in Example 29.

Later, we will show how to determine this analytically.

Horizontal asymptotes can take on a variety of forms. Figure 1.36(a) shows that $f(x) = x/(x^2+1)$ has a horizontal asymptote of $y = 0$, where 0 is approached from both above and below.

Figure 1.36(b) shows that $f(x) = x/\sqrt{x^2+1}$ has two horizontal asymptotes; one at $y = 1$ and the other at $y = -1$.

Figure 1.36(c) shows that $f(x) = (\sin x)/x$ has even more interesting behavior than at just $x = 0$; as x approaches $\pm\infty$, $f(x)$ approaches 0, but oscillates as it does this.

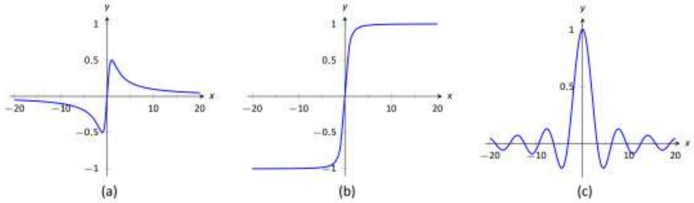


FIGURE 1.36: Considering different types of horizontal asymptotes.

We can analytically evaluate limits at infinity for rational functions once we understand $\lim\limits_{x\to\infty} 1/x$. As x gets larger and larger, the $1/x$ gets smaller and smaller, approaching 0. We can, in fact, make $1/x$ as small as we want by choosing a large enough value of x. Given ϵ, we can make $1/x<\epsilon$ by choosing $x>1/\epsilon$. Thus we have $\lim\limits_{x\to\infty} 1/x=0$.

It is now not much of a jump to conclude the following:

$$\lim_{x\to\infty}\frac{1}{x^n}=0 \quad \text{and} \quad \lim_{x\to-\infty}\frac{1}{x^n}=0 \tag{1.6.6}$$

Now suppose we need to compute the following limit:

$$\lim_{x\to\infty}\frac{x^3+2x+1}{4x^3-2x^2+9}. \tag{1.6.7}$$

A good way of approaching this is to divide through the numerator and denominator by x^3 (hence dividing by 1), which is the largest power of x to appear in the function. Doing this, we get

$$\begin{aligned}\lim_{x\to\infty}\frac{x^3+2x+1}{4x^3-2x^2+9} &= \lim_{x\to\infty}\frac{1/x^3}{1/x^3}\cdot\frac{x^3+2x+1}{4x^3-2x^2+9}\\ &= \lim_{x\to\infty}\frac{x^3/x^3+2x/x^3+1/x^3}{4x^3/x^3-2x^2/x^3+9/x^3}\\ &= \lim_{x\to\infty}\frac{1+2/x^2+1/x^3}{4-2/x+9/x^3}.\end{aligned}$$

Then using the rules for limits (which also hold for limits at infinity), as well as the fact about limits of $1/x^n$, we see that the limit becomes

$$\frac{1+0+0}{4-0+0}=\frac{1}{4}. \tag{1.6.8}$$

This procedure works for any rational function. In fact, it gives us the following theorem.

Theorem 11: Limits of Rational Functions at Infinity

Let $f(x)$ be a rational function of the following form:

$$f(x)=\frac{a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0}{b_mx^m+b_{m-1}x^{m-1}+\cdots+b_1x+b_0}, \tag{1.6.9}$$

where any of the coefficients may be 0 except for a_n and b_m.

1. If $n=m$, then $\lim\limits_{x\to\infty} f(x)=\lim\limits_{x\to-\infty} f(x)=\frac{a_n}{b_m}$.
2. If $n<m$, then $\lim\limits_{x\to\infty} f(x)=\lim\limits_{x\to-\infty} f(x)=0$.
3. If $n>m$, then $\lim\limits_{x\to\infty} f(x)$ and $\lim\limits_{x\to-\infty} f(x)$ are both infinite.

We can see why this is true. If the highest power of x is the same in both the numerator and denominator (i.e. $n=m$), we will be in a situation like the example above, where we will divide by x^n and in the limit all the terms will approach 0 except for a_nx^n/x^n and b_mx^m/x^n. Since $n=m$, this will leave us with the limit a_n/b_m. If $n<m$, then after dividing through by x^m, all the terms in the numerator will approach 0 in the limit, leaving us with $0/b_m$ or 0. If $n>m$, and we try dividing through by x^n, we end up with all the terms in the denominator tending toward 0, while the x^n term in the numerator does not approach 0. This is indicative of some sort of infinite limit.

Intuitively, as x gets very large, all the terms in the numerator are small in comparison to a_nx^n, and likewise all the terms in the denominator are small compared to b_nx^m. If $n=m$, looking only at these two important terms, we have $(a_nx^n)/(b_nx^m)$. This reduces to a_n/b_m. If $n<m$, the function behaves like $a_n/(b_mx^{m-n})$, which tends toward 0. If $n>m$, the function behaves like a_nx^{n-m}/b_m, which will tend to either ∞ or $-\infty$ depending on the values of n, m, a_n, b_m and whether you are looking for $\lim\limits_{x\to\infty} f(x)$ or $\lim\limits_{x\to-\infty} f(x)$.

With care, we can quickly evaluate limits at infinity for a large number of functions by considering the largest powers of x. For instance, consider again $\lim\limits_{x\to\pm\infty} \frac{x}{\sqrt{x^2+1}}$, graphed in Figure ???(b). When x is very large, $x^2+1\approx x^2$. Thus

$$\sqrt{x^2+1}\approx\sqrt{x^2}=|x|, \quad \text{and} \quad \frac{x}{\sqrt{x^2+1}}\approx\frac{x}{|x|}. \tag{1.6.10}$$

This expression is 1 when x is positive and -1 when x is negative. Hence we get asymptotes of $y=1$ and $y=-1$, respectively.

Example 30: Finding a limit of a rational function

Confirm analytically that $y=1$ is the horizontal asymptote of $f(x)=\frac{x^2}{x^2+4}$, as approximated in Example 29.

Solution

Before using Theorem 11, let's use the technique of evaluating limits at infinity of rational functions that led to that theorem. The largest power of x in f is 2, so divide the numerator and denominator of f by x^2, then take limits.

$$\begin{aligned}\lim_{x\to\infty}\frac{x^2}{x^2+4} &= \lim_{x\to\infty}\frac{x^2/x^2}{x^2/x^2+4/x^2}\\ &= \lim_{x\to\infty}\frac{1}{1+4/x^2}\\ &= \frac{1}{1+0}\\ &= 1.\end{aligned}$$

We can also use Theorem 11 directly; in this case $n=m$ so the limit is the ratio of the leading coefficients of the numerator and denominator, i.e., 1/1 = 1.

Example 31: Finding limits of rational functions

Use Theorem 11 to evaluate each of the following limits.

$$1.\ \lim_{x\to-\infty}\frac{x^2+2x-1}{x^3+1} \qquad 3.\ \lim_{x\to\infty}\frac{x^2-1}{3-x} \tag{1.6.11}$$

$$2.\ \lim_{x\to\infty}\frac{x^2+2x-1}{1-x-3x^2} \tag{1.6.12}$$

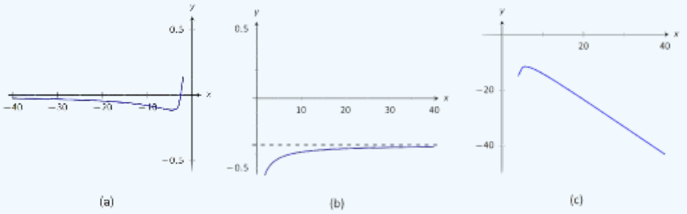


FIGURE 1.37: Visualizing the functions in Example 31.

Solution

1. The highest power of x is in the denominator. Therefore, the limit is 0; see Figure 1.37(a).
2. The highest power of x is x^2, which occurs in both the numerator and denominator. The limit is therefore the ratio of the coefficients of x^2, which is $-1/3$. See Figure 1.37(b).
3. The highest power of x is in the numerator so the limit will be ∞ or $-\infty$. To see which, consider only the dominant terms from the numerator and denominator, which are x^2 and $-x$. The expression in the limit will behave like $x^2/(-x)=-x$ for large values of x. Therefore, the limit is $-\infty$. See Figure 1.37(c).

Chapter Summary

In this chapter we:

- defined the limit,
- found accessible ways to approximate their values numerically and graphically,
- developed a not--so--easy method of proving the value of a limit ($\epsilon - \delta$ proofs),
- explored when limits do not exist,
- defined continuity and explored properties of continuous functions, and
- considered limits that involved infinity.

Why? Mathematics is famous for building on itself and calculus proves to be no exception. In the next chapter we will be interested in "dividing by 0." That is, we will want to divide a quantity by a smaller and smaller number and see what value the quotient approaches. In other words, we will want to find a limit. These limits will enable us to, among other things, determine *exactly* how fast something is moving when we are only given position information.

Later, we will want to add up an infinite list of numbers. We will do so by first adding up a finite list of numbers, then take a limit as the number of things we are adding approaches infinity. Surprisingly, this sum often is finite; that is, we can add up an infinite list of numbers and get, for instance, 42.

These are just two quick examples of why we are interested in limits. Many students dislike this topic when they are first introduced to it, but over time an appreciation is often formed based on the scope of its applicability.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 1.6: Limits Involving Infinity is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

1.E: Applications of Limits (Exercises)

1.1: An Introduction to Limits

Terms and Concepts

1. In your own words, what does it mean to "find the limit of $f(x)$ as x approaches 3"?
2. An expression of the form $\frac{0}{0}$ is called ______.
3. T/F: The limit of $f(x)$ as x approaches 5 is $f(5)$.
4. Describe three situations where $\lim\limits_{x\to c} f(x)$ does not exist.
5. In your own words, what is a difference quotient.

Problems

In Exercises 6-16, approximate the given limits both numerically and graphically.

6. $\lim\limits_{x\to 1} x^2+3x-5$

7. $\lim\limits_{x\to 0} x^3-3x^2+x-5$

8. $\lim\limits_{x\to 0} \frac{x+1}{x^2+3x}$

9. $\lim\limits_{x\to 3} \frac{x^2-2x-3}{x^2-4x+3}$

10. $\lim\limits_{x\to -1} \frac{x^2+8x+7}{x^2+6x+5}$

11. $\lim\limits_{x\to 2} \frac{x^2+7x+10}{x^2-4x+4}$

12. $\lim\limits_{x\to 2}$, where $f(x)=\begin{cases} x+2 & x\leq 2 \\ 3x-5 & x>2 \end{cases}$.

13. $\lim\limits_{x\to 3}$, where $f(x)=\begin{cases} x^2-x+1 & x\leq 3 \\ 2x+1 & x>3 \end{cases}$.

14. $\lim\limits_{x\to 0}$, where $f(x)=\begin{cases} \cos x & x\leq 0 \\ x^2+3x+1 & x>0 \end{cases}$.

15. $\lim\limits_{x\to \pi/2}$, where $f(x)=\begin{cases} \sin x & x\leq \pi/2 \\ \cos x & x>\pi/2 \end{cases}$.

In Exercises 16-24, a function f and a value a are given. Approximate the limit of the difference quotient, $\lim\limits_{h\to 0}\frac{f(a+h)-f(a)}{h}$, using $h=\pm 0.1,\ \pm 0.01$.

16. $f(x)=-7x+2, \quad a=3$

17. $f(x)=9x+0.06, \quad a=-1$

18. $f(x)=x^2+3x-7, \quad a=1$

19. $f(x)=\frac{1}{x+1}, \quad a=2$

20. $f(x)=-4x^2+5x-1, \quad a=-3$

21. $f(x)=\ln x, \quad a=5$

22. $f(x)=\sin x, \quad a=\pi$

23. $f(x)=\cos x, \quad a=\pi$

1.2: Epsilon-Delta Definition of a Limit

Terms and Concepts

1. What is wrong with the following "definition" of a limit?

"The limit of $f(x)$, as x approaches a, is K'' means that given any $\delta>0$ there exists $\epsilon>0$ such that whenever $|f(x)-K|<\epsilon$, we have $|x-a|<\delta$.

2. Which is given first in establishing a limit, the x-tolerance or the y-tolerance?

3. T/F: ϵ must always be positive.

4. T/F: δ must always be positive.

Problems

In Exercises 5-11, prove the given limit using an $\epsilon - \delta$ proof.

5. $\lim\limits_{x\to 5} 3 - x + -2$

6. $\lim\limits_{x\to 3} x^2 - 3 = 6$

7. $\lim\limits_{x\to 4} x^2 + x - 5 = 15$

8. $\lim\limits_{x\to 2} x^3 - 1 = 7$

9. $\lim\limits_{x\to 2} 5 = 5$

10. $\lim\limits_{x\to 0} e^{2x} - 1 = 0$

11. $\lim\limits_{x\to 0} \sin x = 0$ (Hint: use the fact that $|\sin x| \leq |x|$, with equality only when $x = 0$.)

1.3: Finding Limits Analytically

Terms and Concepts

1. Explain in your own words, without using $\varepsilon - \delta$ formality, why $\lim\limits_{x\to c} b = b$.

2. Explain in your own words, without using $\varepsilon - \delta$ formality, why $\lim\limits_{x\to c} x = c$.

3. What does the text mean when it says that certain functions' "behavior is 'nice' in terms of limits"? What, in particular, is "nice"?

4. Sketch a graph that visually demonstrates the Squeeze Theorem.

5. You are given the following information:

    (a) $\lim\limits_{x\to 1} f(x) = 0$

    (b)$\lim\limits_{x\to 1} g(x) = 0$

    (c)$\lim\limits_{x\to 1} f(x)/g(x) = 2$

    What can be said about the relative sizes of $f(x)$ and $g(x)$ as x approaches 1?

Problems

Using:

$$\lim_{x\to 9} f(x) = 6 \qquad \lim_{x\to 6} f(x) = 9 \tag{1.E.1}$$

$$\lim_{x\to 9} g(x) = 3 \qquad \lim_{x\to 6} g(x) = 3 \tag{1.E.2}$$

evaluate the limits given in Exercises 6-13, where possible. If it is not possible to know, state so.

6. $\lim\limits_{x\to 9}(f(x) + g(x))$

7. $\lim\limits_{x\to 9}(3f(x)/g(x))$

8. $\lim\limits_{x\to 9}\left(\frac{f(x)-2g(x)}{g(x)}\right)$

9. $\lim\limits_{x\to 6}\left(\frac{f(x)}{3-g(x)}\right)$

10. $\lim\limits_{x\to 9} g(f(x))$

11. $\lim\limits_{x\to 6} f(g(x))$

12. $\lim\limits_{x\to 6} g(f(f(x)))$

13. $\lim\limits_{x\to 6} f(x)g(x)-f^2(x)+g^2(x)$

Using

$$\lim_{x\to 1} f(x)=2 \qquad \lim_{x\to 10} f(x)=1 \tag{1.E.3}$$

$$\lim_{x\to 1} g(x)=0 \qquad \lim_{x\to 10} g(x)=\pi \tag{1.E.4}$$

evaluate the limits given in Exercises 14-17, where possible. If it is not possible to know, state so.

14. $\lim\limits_{x\to 1} f(x)^{g(x)}$

15. $\lim\limits_{x\to 10} \cos(g(x))$

16. $\lim\limits_{x\to 1} f(x)g(x)$

17. $\lim\limits_{x\to 1} g(5f(x))$

In Exercises 18-32, evaluate the given limit.

18. $\lim\limits_{x\to 3} x^2-3x+7$

19. $\lim\limits_{x\to \pi} \left(\frac{x-3}{x+5}\right)^7$

20. $\lim\limits_{x\to \pi/4} \cos x\sin x$

21. $\lim\limits_{x\to 0} \ln x$

22. $\lim\limits_{x\to 3} 4^{x^3-8x}$

23. $\lim\limits_{x\to \pi/6} \csc x$

24. $\lim\limits_{x\to 0} \ln(1+x)$

25. $\lim\limits_{x\to \pi} \frac{x^2+3x+5}{5x^2-2x-3}$

26.$\lim\limits_{x\to \pi} \frac{3x+1}{1-x}$

27.$\lim\limits_{x\to 6} \frac{x^2-4x-12}{x^2-13x+42}$

28.$\lim\limits_{x\to 0} \frac{x^2+2x}{x^2-2x}$

29.$\lim\limits_{x\to 2} \frac{x^2+6x-16}{x^2-3x+2}$

30.$\lim\limits_{x\to 2} \frac{x^2-5x-14}{x^2+10x+16}$

31. $\lim\limits_{x\to -2} \frac{x^2-5x-14}{x^2+10x+16}$

32. $\lim\limits_{x\to -1} \frac{x^2+9x+8}{x^2-6x-7}\backslash$

Use the Squeeze Theorem in Exercises 33-36, where appropriate, to evaluate the given limit.

33. $\lim\limits_{x\to 0} x\sin\left(\frac{1}{x}\right)$

34. $\lim\limits_{x\to 0} \sin x\cos\left(\frac{1}{x^2}\right)$

35. $\lim\limits_{x\to 1} f(x)$, where $3x-2\leq f(x)\leq x^3$.

36. $\lim\limits_{x\to 3+} f(x)$, where $6x-9\leq f(x)\leq x^2$ on [0,3].

Exercises 37-40, challenge your understanding of limits but can be evaluated using the knowledge gained in this section.

37. $\lim\limits_{x\to 0} \frac{\sin 3x}{x}$

38. $\lim\limits_{x\to 0} \frac{\sin 5x}{8x}$

39. $\lim\limits_{x\to 0} \frac{\ln(1+x)}{x}$

40. $\lim\limits_{x\to 0}\frac{\sin x}{x}$, where x is measured in degrees not radians.

1.4: One Sided Limits

Terms and Concepts

1. What are the three ways in which a limit may fail to exist?

2. T/F: If $\lim\limits_{x\to 1-} f(x)=5$, then $\lim\limits_{x\to 1} f(x)=5$

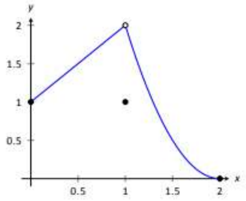
3. T/F: If $\lim\limits_{x\to 1-} f(x)=5$, then $\lim\limits_{x\to 1+} f(x)=5$

4. T/F: If $\lim\limits_{x\to 1} f(x)=5$, then $\lim\limits_{x\to 1-} f(x)=5$

Problems

In Exercises 5-12, evaluate each expression using the given graph of $f(x)$.

5.



(a) $\lim\limits_{x\to 1^-} f(x)$

(b) $\lim\limits_{x\to 1^+} f(x)$

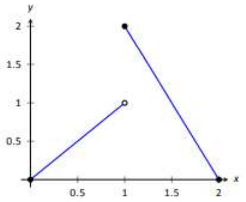
(c) $\lim\limits_{x\to 1} f(x)$

(d) $f(1)$

(e) $\lim\limits_{x\to 0^-} f(x)$

(f) $\lim\limits_{x\to 0^+} f(x)$

6.



(a) $\lim\limits_{x\to 1^-} f(x)$

(b) $\lim\limits_{x\to 1^+} f(x)$

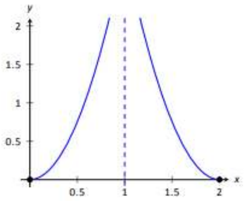
(c) $\lim\limits_{x\to 1} f(x)$

(d) $f(1)$

(e) $\lim\limits_{x\to 2^-} f(x)$

(f) $\lim\limits_{x\to 2^+} f(x)$

7.



(a) $\lim\limits_{x\to 1^-} f(x)$

(b) $\lim\limits_{x\to 1^+} f(x)$

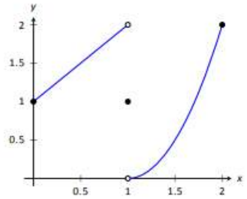
(c) $\lim\limits_{x\to 1} f(x)$

(d) $f(1)$

(e) $\lim\limits_{x\to 2^-} f(x)$

(f) $\lim\limits_{x\to 0^+} f(x)$

8.



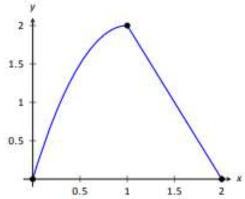
(a) $\lim_{x\to 1^-} f(x)$

(b) $\lim_{x\to 1^+} f(x)$

(c) $\lim_{x\to 1} f(x)$

(d) $f(1)$

9.



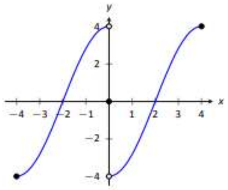
(a) $\lim_{x\to 1^-} f(x)$

(b) $\lim_{x\to 1^+} f(x)$

(c) $\lim_{x\to 1} f(x)$

(d) $f(1)$

10.



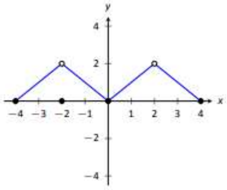
(a) $\lim_{x\to 0^-} f(x)$

(b) $\lim_{x\to 0^+} f(x)$

(c) $\lim_{x\to 0} f(x)$

(d) $f(0)$

11.



(a) $\lim_{x\to -2^-} f(x)$

(b) $\lim_{x\to -2^+} f(x)$

(c) $\lim_{x\to -2} f(x)$

(d) $f(-2)$

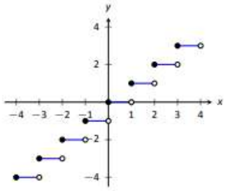
(e) $\lim_{x\to 2^-} f(x)$

(f) $\lim_{x\to 2^+} f(x)$

(g) $\lim_{x\to 2} f(x)$

(h) $f(2)$

12.



Let $-3 \leq a \leq 3$ be an integer.

(a) $\lim\limits_{x\to a^-} f(x)$

(b) $\lim\limits_{x\to a^+} f(x)$

(c) $\lim\limits_{x\to a} f(x)$

(d) $f(a)$

In Exercises 13-21, evaluate the given limits of the piecewise defined functions f.

13. $f(x) = \begin{cases} x+1 & x \leq 1 \\ x^2-5 & x > 1 \end{cases}$

(a) $\lim\limits_{x\to 1^-} f(x)$

(b) $\lim\limits_{x\to 0^+} f(x)$

(c) $\lim\limits_{x\to 1} f(x)$

(d) $f(1)$

14. $f(x) = \begin{cases} 2x^2+5x-1 & x < 0 \\ \sin x & x \geq 0 \end{cases}$

(a) $\lim\limits_{x\to 0^-} f(x)$

(b) $\lim\limits_{x\to 0^+} f(x)$

(c) $\lim\limits_{x\to 0} f(x)$

(d) $f(0)$

15. $f(x) = \begin{cases} x^{-}1 & x < -1 \\ x^3+1 & -1 \leq x \leq 1 \\ x^2+1 & x > 1 \end{cases}$

(a) $\lim\limits_{x\to -1^-} f(x)$

(b) $\lim\limits_{x\to 1^+} f(x)$

(c) $\lim\limits_{x\to -1} f(x)$

(d) $f(-1)$

(e) $\lim\limits_{x\to 1^-} f(x)$

(f) $\lim\limits_{x\to 1^+} f(x)$

(g) $\lim\limits_{x\to 1} f(x)$

(h) $f(1)$

16. $f(x) = \begin{cases} \cos x & x < \pi \\ \sin x & x \geq \pi \end{cases}$

(a) $\lim\limits_{x\to \pi^-} f(x)$

(b) $\lim\limits_{x\to \pi^+} f(x)$

(c) $\lim\limits_{x\to \pi} f(x)$

(d) $f(\pi)$

17. $f(x) = \begin{cases} 1-\cos^2 x & x < a \\ \sin^2 x & x \geq a \end{cases}$, where a is a real number.

(a) $\lim\limits_{x\to a^-} f(x)$

(b) $\lim\limits_{x\to a^+} f(x)$

(c) $\lim\limits_{x\to a} f(x)$

(d) $f(a)$

18. $f(x) = \begin{cases} x+1 & x<1 \\ 1 & x=1 \\ x-1 & x>1 \end{cases}$

(a) $\lim\limits_{x\to 1^-} f(x)$

(b) $\lim\limits_{x\to 1^+} f(x)$

(c) $\lim\limits_{x\to 1} f(x)$

(d) $f(1)$

19. $f(x) = \begin{cases} x^2 & x<2 \\ x+1 & x=2 \\ -x^2+2x+4 & x>2 \end{cases}$

(a) $\lim\limits_{x\to 2^-} f(x)$

(b) $\lim\limits_{x\to 2^+} f(x)$

(c) $\lim\limits_{x\to 2} f(x)$

(d) $f(2)$

20. $f(x) = \begin{cases} a(x-b)^2+c & x<b \\ a(x-b)+c & x\geq b \end{cases}$, where a, b and c are real numbers.

(a) $\lim\limits_{x\to b^-} f(x)$

(b) $\lim\limits_{x\to b^+} f(x)$

(c) $\lim\limits_{x\to b} f(x)$

(d) $f(b)$

21. $f(x) = \begin{cases} \frac{|x|}{x} & x\neq 0 \\ 0 & x=0 \end{cases}$

(a) $\lim\limits_{x\to 0^-} f(x)$

(b) $\lim\limits_{x\to 0^+} f(x)$

(c) $\lim\limits_{x\to 0} f(x)$

(d) $f(0)$

Review

22. Evaluate the limit: $\lim\limits_{x\to -1} \frac{x^2+5x+4}{x^2-3x-4}$

23. Evaluate the limit: $\lim\limits_{x\to -4} \frac{x^2-16}{x^2-4x-32}$

24. Evaluate the limit: $\lim\limits_{x\to -6} \frac{x^2-15x+54}{x^2-6x}$

25. Approximate the limit numerically: $\lim\limits_{x\to 0.4} \frac{x^2-4.4x+1.6}{x^2-0.4x}$

26. Approximate the limit numerically: $\lim\limits_{x\to 0.2} \frac{x^2+5.8x-1.2}{x^2-4.2x+0.8}$

1.5: Continuity

Terms and Concepts

1. In your own words, describe what it means for a function to be continuous.

2. In your own words, describe what the Intermediate Value Theorem states.

3. What is a “root” of a function?

4. Given functions f and g on an interval I, how can the Bisection Method be used to find a value c where $f(c)=g(c)$?

5. T/F: If f is defined on an open interval containing c, and $\lim\limits_{x\to c} f(x)$ exists, then f is continuous at c.

6. T/F: If f is continuous at c, then $\lim\limits_{x\to c} f(x)$ exists

7. T/F: If f is continuous at c, then $\lim\limits_{x\to c^+} f(x) = f(c)$.

8. T/F: If f is continuous on [a, b], then $\lim\limits_{x\to a^-} f(x) = f(a)$.

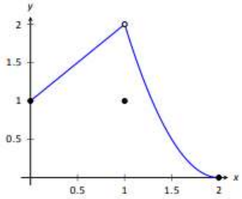
9. T/F: If f is continuous on [0, 1) and [1, 2), then f is continuous on [0, 2).

10. T/F: The sum of continuous functions is also continuous.

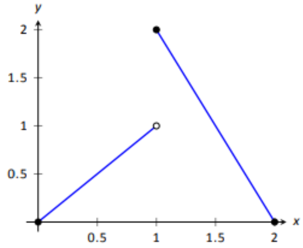
Problems

In Exercises 11-17, a graph of a function f is given along with a value a. Determine if f is continuous at a; if it is not, state why it is not.

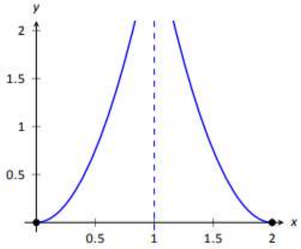
11. $a = 1$



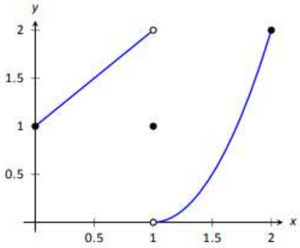
12. $a = 1$



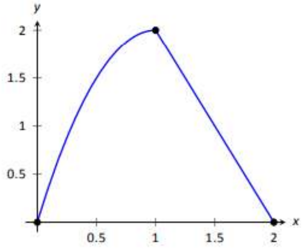
13. $a = 1$



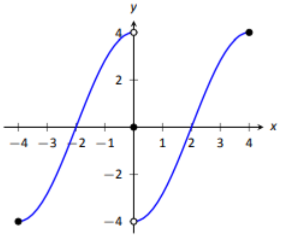
14. $a = 0$



15. $a = 1$



16. $a = 4$

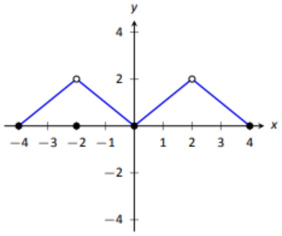


17.

(a) $a = -2$

(b) $a = 0$

(c) $a = 2$



In Exercises 18-21, determine if f is continuous at the indicated values. If not, explain why.

18. $f(x) = \begin{cases} 1 & x = 0 \\ \frac{\sin x}{x} & x > 0 \end{cases}$

(a) $x = 0$

(b) $x = \pi$

19. $f(x) = \begin{cases} x^3 - x & x < 1 \\ x - 2 & x \geq 1 \end{cases}$

(a) $x = 0$

(b) $x = 1$

20. $f(x) = \begin{cases} \frac{x^2+5x+4}{x^2+3x+2} & x \neq -1 \\ 3 & x = -1 \end{cases}$

(a) $x = -1$

(b) $x = 10$

21. $f(x) = \begin{cases} \frac{x^2-64}{x^2-11x+24} & x \neq 8 \\ 5 & x = 8 \end{cases}$

(a) $x = 0$

(b) $x = 8$

In Exercises 22-32, give the intervals on which the given function is continuous.

22. $f(x) = x^2 - 3x + 9$

23. $g(x) = \sqrt{x^2 - 4}$

24. $h(k) = \sqrt{1-k} + \sqrt{k+1}$

25. $f(t) = \sqrt{5t^2 - 30}$

26. $g(t) = \frac{1}{\sqrt{1-t^2}}$

27. $g(x) = \frac{1}{1+x^2}$

28. $f(x) = e^x$

29. $g(s) = \ln s$

30. $h(t) = \cos t$

31. $f(k) = \sqrt{1-e^k}$

32. $f(x) = \sin(e^x + x^2)$

33. Let f be continuous on [1,5] where $f(1) = -2$ and $f(5) = -10$. Does a value $1 < c < 5$ exist such that $f(c) = -9$? Why/why not?

34. Let g be continuous on [-3,7] where $g(0) = 0$ and $g(2) = 25$. Does a value $-3 < c < 7$ exist such that $g(c) = 15$? Why/why not?

35. Let f be continuous on [-1,1] where $f(-1) = -10$ and $f(1) = 10$. Does a value $-1 < c < 1$ exist such that $f(c) = 11$? Why/why not?

36. Let h be continuous on [-1,1] where $h(-1) = -10$ and $h(1) = 10$. Does a value $-1 < c < 1$ exist such that $h(c) = 0$? Why/why not?

In Exercises 37-40, use the Bisection Method to approximate, accurate to two decimal places, the value of the root of the given function in the given interval.

37. $f(x) = x^2 + 2x - 4$ on $[1, 1.5]$.

38. $f(x) = \sin x - 1/2$ on $[0.5, 0.55]$.

39. $f(x) = e^x - 2$ on $[0.65, 0.7]$.

40. $f(x) = \cos x - \sin x$ on $[0.7, 0.8]$.

Review

41. Let $f(x) = \begin{cases} x^2 - 5 & x < 5 \\ 5x & x \geq 5 \end{cases}$.

(a) $\lim\limits_{x\to 5^-} f(x)$

(b) $\lim\limits_{x\to 5^+} f(x)$

(c) $\lim\limits_{x\to 5} f(x)$

(d) $f(5)$

42. Numerically approximate the following limits:

(a) $\lim\limits_{x\to 4/5^+} \frac{x^2-8.2x-7.2}{x^2+5.8x+4}$

(b) $\lim\limits_{x\to 4/5^-} \frac{x^2-8.2x-7.2}{x^2+5.8x+4}$

43. Give an example of function $f(x)$ for which $\lim\limits_{x\to 0} f(x)$ does not exist.

1.6: Limits Involving Infinity

Terms and Concepts

1. T/F: If $\lim\limits_{x\to 5} f(x) = \infty$, then we are implicitly stating that the limit exists.

2. T/F: If $\lim\limits_{x\to \infty} f(x) = 5$, then we are implicitly stating that the limit exists.

3. T/F: If $\lim\limits_{x\to 1^-} f(x) = -\infty$, then $\lim\limits_{x\to 1^+} f(x) = \infty$.

4. T/F: If $\lim\limits_{x\to 5} f(x) = \infty$, then f has a vertical asymptote at $x = 5$.

5. T/F: $\infty/0$ is not an indeterminate form.

6. List 5 indeterminate forms.

7. Construct a function with a vertical asymptote at x = 5 and a horizontal asymptote at y = 5.

8. Let $\lim\limits_{x\to 7} f(x) = \infty$. Explain how we know that f is/is not continuous at $x = 7$.

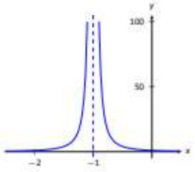
Problems

In Exercises 9-14, evaluate the given limits using the graph of the function.

9. $f(x) = \frac{1}{(x+1)^2}$

(a) $\lim\limits_{x\to -1^-} f(x)$

(b) $\lim\limits_{x\to -1^+} f(x)$



10. $f(x)=\frac{1}{(x-3)(x-5)^2}$

(a) $\lim\limits_{x\to 3^-} f(x)$

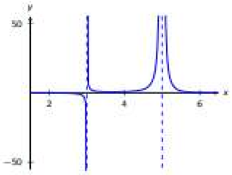
(b) $\lim\limits_{x\to 3^+} f(x)$

(c) $\lim\limits_{x\to 3} f(x)$

(d) $\lim\limits_{x\to 5^-} f(x)$

(e) $\lim\limits_{x\to 5^+} f(x)$

(f) $\lim\limits_{x\to 5} f(x)$



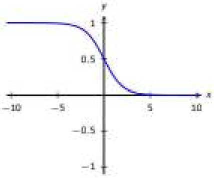
11. $f(x)=\frac{1}{e^x+1}$

(a) $\lim\limits_{x\to -\infty} f(x)$

(b) $\lim\limits_{x\to \infty} f(x)$

(c) $\lim\limits_{x\to 0^-} f(x)$

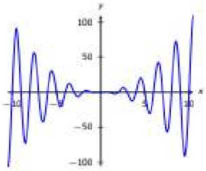
(d) $\lim\limits_{x\to 0^+} f(x)$



12. $f(x)=x^2\sin(\pi x)$

(a) $\lim\limits_{x\to -\infty} f(x)$

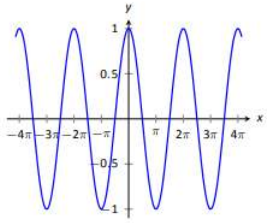
(b) $\lim\limits_{x\to \infty} f(x)$



13. $f(x)=\cos(x)$

(a) $\lim\limits_{x\to -\infty} f(x)$

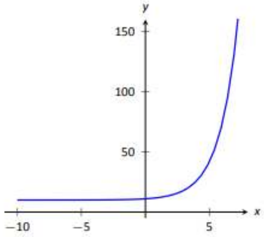
(b) $\lim_{x\to\infty} f(x)$



14. $f(x) = 2^x + 10$

(a) $\lim_{x\to-\infty} f(x)$

(b) $\lim_{x\to\infty} f(x)$



In Exercises 15-18, numerically approximate the following limits:

(a) $\lim_{x\to 3^-} f(x)$

(b) $\lim_{x\to 3^+} f(x)$

(c) $\lim_{x\to 3} f(x)$

15. $f(x) = \frac{x^2-1}{x^2-x-6}$

16. $f(x) = \frac{x^2+5x-36}{x^3-5x^2+3x+9}$

17. $f(x) = \frac{x^2-11x+30}{x^3-4x^2-3x+18}$

18. $f(x) = \frac{x^2-9x+18}{x^2-x-6}$

In Exercises 19-24, identify the horizontal and vertical asymptotes, if any, of the given function.

19. $f(x) = \frac{2x^2-2x-4}{x^2+x-20}$

20. $f(x) = \frac{-3x^2-9x-6}{5x^2-10x-15}$

21. $f(x) = \frac{x^2+2-12}{7x^3-14x^2-21x}$

22. $f(x) = \frac{x^2-9}{9x-9}$

23. $f(x) = \frac{x^2-9}{9x+27}$

24. $f(x) = \frac{x^2-1}{-x^2-1}$

In Exercises 25-28, evaluate the given limit.

25. $\lim_{x\to\infty} \frac{x^3+2x^2+1}{x-5}$

26. $\lim_{x\to\infty} \frac{x^3+2x^2+1}{5-x}$

27. $\lim_{x\to\infty} \frac{x^3+2x^2+1}{x^2-5}$

28. $\lim_{x\to\infty} \frac{x^3+2x^2+1}{5-x^2}$

Review

29. Use an $\varepsilon-\delta$ proof to show that $\lim_{x\to 1} 5x-2=3$.

30. Let $\lim\limits_{x\to 2} f(x) = 3$ and $\lim\limits_{x\to 2} g(x) = -1$. Evaluate the following limits.

(a) $\lim\limits_{x\to 2}(f+g)(x)$

(b) $\lim\limits_{x\to 2}(fg)(x)$

(c) $\lim\limits_{x\to 2}(f/g)(x)$

(d) $\lim\limits_{x\to 2} f(x)^{g(x)}$

31. Let $f(x) = \begin{cases} x^2-1 & x<3 \\ x+5 & x\geq 3 \end{cases}$. Is f continuous everywhere?

32. Evaluate the limit: $\lim\limits_{x\to c} \ln x$.

1.E: Applications of Limits (Exercises) is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

2: Derivatives

The previous chapter introduced the most fundamental of calculus topics: the limit. This chapter introduces the second most fundamental of calculus topics: the derivative. Limits describe where a function is going; derivatives describe how fast the function is going.

Topic hierarchy

2.1: Instantaneous Rates of Change- The Derivative
2.2: Interpretations of the Derivative
2.3: Basic Differentiation Rules
2.4: The Product and Quotient Rules
2.5: The Chain Rule
2.6: Implicit Differentiation
2.7: Derivatives of Inverse Functions
2.E: Applications of Derivatives(Exercises)

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 2: Derivatives is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

2.1: Instantaneous Rates of Change- The Derivative

A common amusement park ride lifts riders to a height then allows them to freefall a certain distance before safely stopping them. Suppose such a ride drops riders from a height of 150 feet. Students of physics may recall that the height (in feet) of the riders, t seconds after freefall (and ignoring air resistance, etc.) can be accurately modeled by $f(t) = -16t^2 + 150$.

Using this formula, it is easy to verify that, without intervention, the riders will hit the ground at $t = 2.5\sqrt{1.5} \approx 3.06$ seconds. Suppose the designers of the ride decide to begin slowing the riders' fall after 2 seconds (corresponding to a height of 86 ft.). How fast will the riders be traveling at that time?

We have been given a *position* function, but what we want to compute is a velocity at a specific point in time, i.e., we want an *instantaneous velocity*. We do not currently know how to calculate this.

However, we do know from common experience how to calculate an *average velocity*. (If we travel 60 miles in 2 hours, we know we had an average velocity of 30 mph.) We looked at this concept in Section 1.1 when we introduced the difference quotient. We have

$$\frac{\text{change in distance}}{\text{change in time}} = \frac{\text{"rise"}}{\text{run}} = \text{average velocity}. \tag{2.1.1}$$

We can approximate the instantaneous velocity at $t = 2$ by considering the average velocity over some time period containing $t = 2$. If we make the time interval small, we will get a good approximation. (This fact is commonly used. For instance, high speed cameras are used to track fast moving objects. Distances are measured over a fixed number of frames to generate an accurate approximation of the velocity.)

Consider the interval from $t = 2$ to $t = 3$ (just before the riders hit the ground). On that interval, the average velocity is

$$\frac{f(3) - f(2)}{3 - 2} = \frac{f(3) - f(2)}{1} = -80 \text{ ft/s}, \tag{2.1.2}$$

where the minus sign indicates that the riders are moving *down*. By narrowing the interval we consider, we will likely get a better approximation of the instantaneous velocity. On $[2, 2.5]$ we have

$$\frac{f(2.5) - f(2)}{2.5 - 2} = \frac{f(2.5) - f(2)}{0.5} = -72 \text{ ft/s}. \tag{2.1.3}$$

We can do this for smaller and smaller intervals of time. For instance, over a time span of $1/10^{\text{th}}$ of a second, i.e., on $[2, 2.1]$, we have

$$\frac{f(2.1) - f(2)}{2.1 - 2} = \frac{f(2.1) - f(2)}{0.1} = -65.6 \text{ ft/s}. \tag{2.1.4}$$

Over a time span of $1/100^{\text{th}}$ of a second, on $[2, 2.01]$, the average velocity is

$$\frac{f(2.01) - f(2)}{2.01 - 2} = \frac{f(2.01) - f(2)}{0.01} = -64.16 \text{ ft/s}. \tag{2.1.5}$$

What we are really computing is the average velocity on the interval $[2, 2+h]$ for small values of h. That is, we are computing

$$\frac{f(2+h) - f(2)}{h} \tag{2.1.6}$$

where h is small.

What we really want is for $h = 0$, but this, of course, returns the familiar "$0/0$" indeterminate form. So we employ a limit, as we did in Section 1.1.

We can approximate the value of this limit numerically with small values of h as seen in Figure 2.1. It looks as though the velocity is approaching -64 ft/s. Computing the limit directly gives

$$\begin{aligned}\lim_{h\to 0}\frac{f(2+h)-f(2)}{h} &= \lim_{h\to 0}\frac{-16(2+h)^2+150-(-16(2)^2+150)}{h}\\ &= \lim_{h\to 0}\frac{-64h-16h^2}{h}\\ &= \lim_{h\to 0}-64-16h\\ &= -64.\end{aligned}$$

h	Average Velocity ft/s
1	-80
0.5	-72
0.1	-65.6
0.01	-64.16
0.001	-64.016

Figure 2.1: Approximating the instantaneous velocity with average velocities over a small time period h.

Graphically, we can view the average velocities we computed numerically as the slopes of secant lines on the graph of f going through the points $(2, f(2))$ and $(2+h, f(2+h))$. In Figure 2.2, the secant line corresponding to $h=1$ is shown in three contexts. Figure 2.2(a) shows a "zoomed out" version of f with its secant line. In (b), we zoom in around the points of intersection between f and the secant line. Notice how well this secant line approximates f between those two points -- it is a common practice to approximate functions with straight lines.

As $h \to 0$, these secant lines approach the *tangent line*, a line that goes through the point $(2, f(2))$ with the special slope of -64. In parts (c) and (d) of Figure 2.2, we zoom in around the point $(2, 86)$. In (c) we see the secant line, which approximates f well, but not as well the tangent line shown in (d).

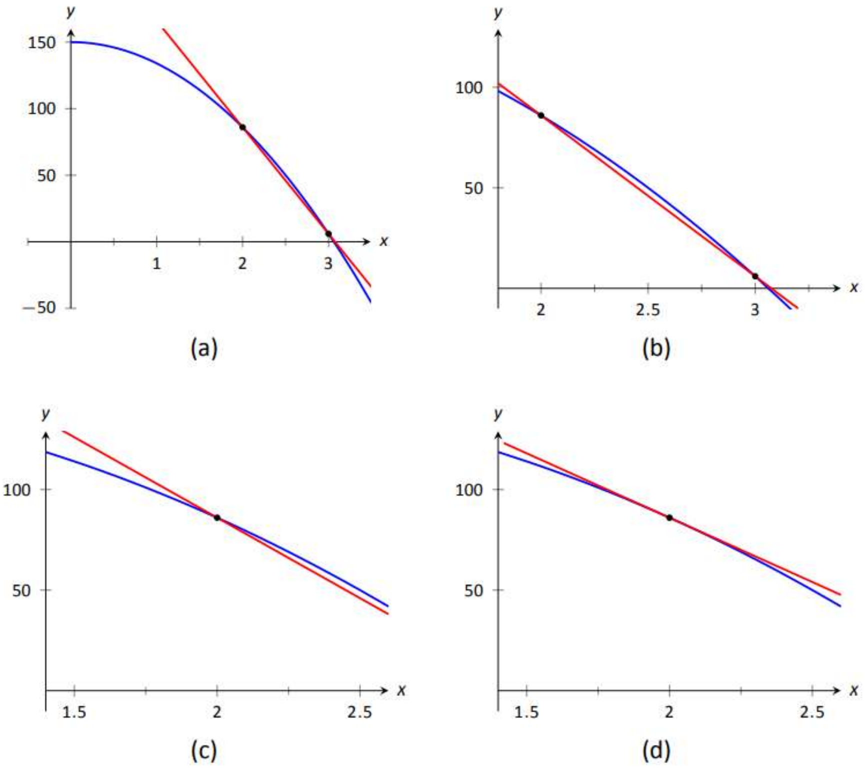


Figure 2.2: Parts (a), (b) and (c) show the secant line to $f(x)$ with $h=1$, zoomed in different amounts. Part (d) shows the tangent line to f at $x=2$.

We have just introduced a number of important concepts that we will flesh out more within this section. First, we formally define two of them.

Definition 7: Derivative at a Point

Let f be a continuous function on an open interval I and let c be in I. The **derivative of f at c**, denoted $f'(c)$, is

$$\lim_{h\to 0} \frac{f(c+h)-f(c)}{h}, \tag{2.1.7}$$

provided the limit exists. If the limit exists, we say that f **is differentiable at** c}; if the limit does not exist, then f **is not differentiable at** c}. If f is differentiable at every point in I, then f **is differentiable on** I.

Definition 8: Tangent Line

Let f be continuous on an open interval I and differentiable at c, for some c in I. The line with equation $\ell(x) = f'(c)(x - c) + f(c)$ is the **tangent line** to the graph of f at c; that is, it is the line through $(c, f(c))$ whose slope is the derivative of f at c.

Some examples will help us understand these definitions.

Example 32: Finding derivatives and tangent lines

Let $f(x) = 3x^2 + 5x - 7$. Find:

1. $f'(1)$
2. The equation of the tangent line to the graph of f at $x = 1$.
3. $f'(3)$
4. The equation of the tangent line to the graph f at $x = 3$.

Solution

1. We compute this directly using Definition 7.

$$\begin{aligned} f'(1) &= \lim_{h\to 0} \frac{f(1+h)-f(1)}{h} \\ &= \lim_{h\to 0} \frac{3(1+h)^2+5(1+h)-7-(3(1)^2+5(1)-7)}{h} \\ &= \lim_{h\to 0} \frac{3h^2+11h}{h} \\ &= \lim_{h\to 0} 3h+11 = 11. \end{aligned}$$

2. The tangent line at $x = 1$ has slope $f'(1)$ and goes through the point $(1, f(1)) = (1, 1)$. Thus the tangent line has equation, in point-slope form, $y = 11(x - 1) + 1$. In slope-intercept form we have $y = 11x - 10$.
3. Again, using the definition,

$$\begin{aligned} f'(3) &= \lim_{h\to 0} \frac{f(3+h)-f(3)}{h} \\ &= \lim_{h\to 0} \frac{3(3+h)^2+5(3+h)-7-(3(3)^2+5(3)-7)}{h} \\ &= \lim_{h\to 0} \frac{3h^2+23h}{h} \\ &= \lim_{h\to 0} 3h+23 \\ &= 23. \end{aligned}$$

4. The tangent line at $x = 3$ has slope 23 and goes through the point $(3, f(3)) = (3, 35)$. Thus the tangent line has equation $y = 23(x - 3) + 35 = 23x - 34$.

   A graph of $f(x) = 3x^2 + 5x - 7$ and its tangent lines at $x = 1$ and $x = 3$.

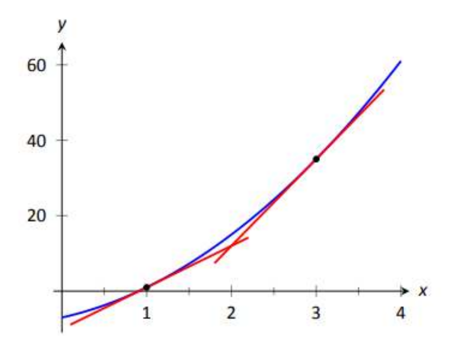


Figure 2.3: A graph of $f(x)=3x^2+5x-7$ and its tangent lines at $x=1$ and $x=3$.

Another important line that can be created using information from the derivative is the **normal line**. It is perpendicular to the tangent line, hence its slope is the opposite--reciprocal of the tangent line's slope.

Definition 9: Normal Line

Let f be continuous on an open interval I and differentiable at c, for some c in I. The **normal line** to the graph of f at c is the line with equation

$$n(x)=\frac{-1}{f'(c)}(x-c)+f(c), \tag{2.1.8}$$

where $f'(c)\neq 0$. When $f'(c)=0$, the normal line is the vertical line through $\big(c,f(c)\big)$; that is, $x=c$.

Example 33: Finding equations of normal lines

Let $f(x)=3x^2+5x-7$, as in Example 32. Find the equations of the normal lines to the graph of f at $x=1$ and $x=3$.

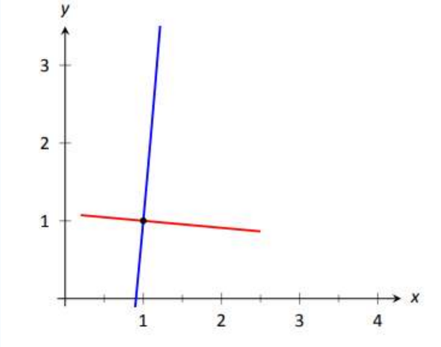


Figure 2.4: A graph of $f(x)=3x^2+5x-7$, along with its normal line at $x=1$.

Solution

In Example 32, we found that $f'(1)=11$. Hence at $x=1$, the normal line will have slope $-1/11$. An equation for the normal line is

$$n(x)=\frac{-1}{11}(x-1)+1. \tag{2.1.9}$$

The normal line is plotted with $y=f(x)$ in Figure 2.4. Note how the line looks perpendicular to f. (A key word here is "looks." Mathematically, we say that the normal line *is* perpendicular to f at $x=1$ as the slope of the normal line is the opposite--reciprocal of the slope of the tangent line. However, normal lines may not always *look* perpendicular. The aspect ratio of the picture of the graph plays a big role in this.)

We also found that $f'(3)=23$, so the normal line to the graph of f at $x=3$ will have slope $-1/23$. An equation for the normal line is

$$n(x)=\frac{-1}{23}(x-3)+35. \tag{2.1.10}$$

Linear functions are easy to work with; many functions that arise in the course of solving real problems are not easy to work with. A common practice in mathematical problem solving is to approximate difficult functions with not--so--difficult functions. Lines are a common choice. It turns out that at any given point on the graph of a differentiable function f, the best linear approximation to f is its tangent line. That is one reason we'll spend considerable time finding tangent lines to functions.

One type of function that does not benefit from a tangent--line approximation is a line; it is rather simple to recognize that the tangent line to a line is the line itself. We look at this in the following example.

Example 34: Finding the Derivative of a Line

Consider $f(x)=3x+5$. Find the equation of the tangent line to f at $x=1$ and $x=7$.

Solution

We find the slope of the tangent line by using Definition 7.

$$\begin{aligned} f'(1) &= \lim_{h\to 0}\frac{f(1+h)-f(1)}{h} \\ &= \lim_{h\to 0}\frac{3(1+h)+5-(3+5)}{h} \\ &= \lim_{h\to 0}\frac{3h}{h} \\ &= \lim_{h\to 0} 3 \\ &= 3. \end{aligned}$$

We just found that $f'(1)=3$. That is, we found the *instantaneous rate of change* of $f(x)=3x+5$ is 3. This is not surprising; lines are characterized by being the *only* functions with a *constant rate of change*. That rate of change is called the *slope* of the line. Since their rates of change are constant, their *instantaneous* rates of change are always the same; they are all the slope.

So given a line $f(x)=ax+b$, the derivative at any point x will be a; that is, $f'(x)=a$.

It is now easy to see that the tangent line to the graph of f at $x=1$ is just f, with the same being true for $x=7$.

We often desire to find the tangent line to the graph of a function without knowing the actual derivative of the function. In these cases, the best we may be able to do is approximate the tangent line. We demonstrate this in the next example.

Example 35: Numerical Approximation of the Tangent Line

Approximate the equation of the tangent line to the graph of $f(x)=\sin x$ at $x=0$.

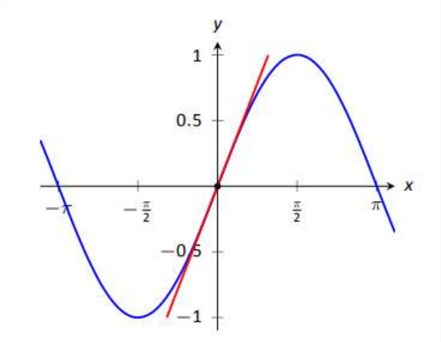


Figure 2.5: $f(x)=\sin x$ graphed with an approximation to its tangent line at $x=0$.

Solution

In order to find the equation of the tangent line, we need a slope and a point. The point is given to us: $(0,\sin 0)=(0,0)$. To compute the slope, we need the derivative. This is where we will make an approximation. Recall that

$$f'(0)\approx\frac{\sin(0+h)-\sin 0}{h} \tag{2.1.11}$$

for a small value of h. We choose (somewhat arbitrarily) to let $h=0.1$. Thus

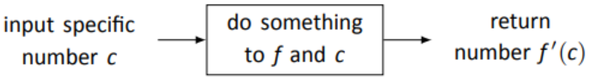
$$f'(0)\approx\frac{\sin(0.1)-\sin 0}{0.1}\approx 0.9983. \tag{2.1.12}$$

0.1

Thus our approximation of the equation of the tangent line is $y=0.9983(x-0)+0=0.9983x$; it is graphed in Figure 2.5. The graph seems to imply the approximation is rather good.

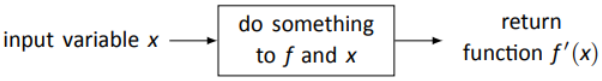
Recall from Section 1.3 that $\lim_{x\to 0}\frac{\sin x}{x}=1$, meaning for values of x near 0, $\sin x\approx x$. Since the slope of the line $y=x$ is 1 at $x=0$, it should seem reasonable that "the slope of $f(x)=\sin x$" is near 1 at $x=0$. In fact, since we *approximated* the value of the slope to be 0.9983, we might guess the *actual value* is 1. We'll come back to this later.

Consider again Example 32. To find the derivative of f at $x=1$, we needed to evaluate a limit. To find the derivative of f at $x=3$, we needed to again evaluate a limit. We have this process:



This process describes a *function*; given one input (the value of c), we return exactly one output (the value of $f'(c)$). The "do something" box is where the tedious work (taking limits) of this function occurs.

Instead of applying this function repeatedly for different values of c, let us apply it just once to the variable x. We then take a limit just once. The process now looks like:



The output is the "derivative function," $f'(x)$. The $f'(x)$ function will take a number c as input and return the derivative of f at c. This calls for a definition.

Definition 10: Derivative Function

Let f be a differentiable function on an open interval I. The function

$$f'(x)=\lim_{h\to 0}\frac{f(x+h)-f(x)}{h} \tag{2.1.13}$$

is **the derivative of** f.

Notation:

Let $y=f(x)$. The following notations all represent the derivative:

$$f'(x)\;=\;y'\;=\;\frac{dy}{dx}\;=\;\frac{df}{dx}\;=\;\frac{d}{dx}(f)\;=\;\frac{d}{dx}(y). \tag{2.1.14}$$

Important: The notation $\frac{dy}{dx}$ is one symbol; it is **not** the fraction "dy/dx". The notation, while somewhat confusing at first, was chosen with care. A fraction--looking symbol was chosen because the derivative has many fraction-like properties. Among other places, we see these properties at work when we talk about the units of the derivative, when we discuss the Chain Rule, and when we learn about integration (topics that appear in later sections and chapters).

Examples will help us understand this definition.

Example 36: Finding the derivative of a function

Let $f(x)=3x^2+5x-7$ as in Example 32. Find $f'(x)$.}

Solution: We apply Definition 10.

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} \\ &= \lim_{h\to 0} \frac{3(x+h)^2+5(x+h)-7-(3x^2+5x-7)}{h} \\ &= \lim_{h\to 0} \frac{3h^2+6xh+5h}{h} \\ &= \lim_{h\to 0} 3h+6x+5 \\ &= 6x+5 \end{aligned}$$

So $f'(x)=6x+5$. Recall earlier we found that $f'(1)=11$ and $f'(3)=23$. Note our new computation of $f'(x)$ affirm these facts.

Example 37: Finding the derivative of a function

Let $f(x)=\frac{1}{x+1}$. Find $f'(x)$.

Solution: We apply Definition 10.

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} \\ &= \lim_{h\to 0} \frac{\frac{1}{x+h+1}-\frac{1}{x+1}}{h} \end{aligned} \tag{2.1.15}$$

Now find common denominator then subtract; pull $1/h$ out front to facilitate reading.

$$\begin{aligned} &= \lim_{h\to 0} \frac{1}{h}\cdot\left(\frac{x+1}{(x+1)(x+h+1)}-\frac{x+h+1}{(x+1)(x+h+1)}\right) \\ &= \lim_{h\to 0} \frac{1}{h}\cdot\left(\frac{x+1-(x+h+1)}{(x+1)(x+h+1)}\right) \\ &= \lim_{h\to 0} \frac{1}{h}\cdot\left(\frac{-h}{(x+1)(x+h+1)}\right) \\ &= \lim_{h\to 0} \frac{-1}{(x+1)(x+h+1)} \\ &= \frac{-1}{(x+1)(x+1)} \\ &= \frac{-1}{(x+1)^2} \end{aligned}$$

So $f'(x)=\frac{-1}{(x+1)^2}$. To practice using our notation, we could also state

$$\frac{d}{dx}\left(\frac{1}{x+1}\right)=\frac{-1}{(x+1)^2}. \tag{2.1.16}$$

Example 38: Finding the derivative of a function

Find the derivative of $f(x)=\sin x$.}

Solution

Before applying Definition 10, note that once this is found, we can find the actual tangent line to $f(x)=\sin x$ at $x=0$, whereas we settled for an approximation in Example 35.

$$\begin{aligned}
f'(x) &= \lim_{h\to0}\frac{\sin(x+h)-\sin x}{h} && \left(\begin{array}{c}\text{Use trig identity}\\ \sin(x+h)=\sin x\cos h+\cos x\sin h\end{array}\right)\\
&= \lim_{h\to0}\frac{\sin x\cos h+\cos x\sin h-\sin x}{h} && \text{(regroup)}\\
&= \lim_{h\to0}\frac{\sin x(\cos h-1)+\cos x\sin h}{h} && \text{(split into two fractions)}\\
&= \lim_{h\to0}\left(\frac{\sin x(\cos h-1)}{h}+\frac{\cos x\sin h}{h}\right) && \left(\text{use }\lim_{h\to0}\frac{\cos h-1}{h}=0\text{ and }\lim_{h\to0}\frac{\sin h}{h}=1\right)\\
&= \sin x\cdot 0+\cos x\cdot 1\\
&= \cos x\,!
\end{aligned}$$

We have found that when $f(x)=\sin x$, $f'(x)=\cos x$. This should be somewhat surprising; the result of a tedious limit process and the sine function is a nice function. Then again, perhaps this is not entirely surprising. The sine function is periodic -- it repeats itself on regular intervals. Therefore its rate of change also repeats itself on the same regular intervals. We should have known the derivative would be periodic; we now know exactly which periodic function it is.

Thinking back to Example 35, we can find the slope of the tangent line to $f(x)=\sin x$ at $x=0$ using our derivative. We approximated the slope as 0.9983; we now know the slope is *exactly* $\cos 0=1$.

Example 39: Finding the derivative of a piecewise defined function

Find the derivative of the absolute value function,

$$f(x)=|x|=\begin{cases}-x & x<0\\ x & x\geq 0\end{cases}. \tag{2.1.17}$$

See Figure 2.6.

Solution

We need to evaluate $\lim_{h\to0}\frac{f(x+h)-f(x)}{h}$. As f is piecewise--defined, we need to consider separately the limits when $x<0$ and when $x>0$.

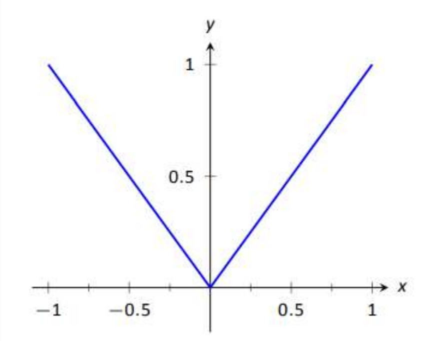


Figure 2.6: The absolute value function, $f(x)=|x|$. Notice how the slope of the lines (and hence the tangent lines) abruptly changes at $x=0$.

When $x<0$:

$$\begin{aligned}
\frac{d}{dx}\big(-x\big) &= \lim_{h\to0}\frac{-(x+h)-(-x)}{h}\\
&= \lim_{h\to0}\frac{-h}{h}\\
&= \lim_{h\to0}-1\\
&= -1.
\end{aligned}$$

When $x>0$, a similar computation shows that $\frac{d}{dx}(x)=1$.

We need to also find the derivative at $x = 0$. By the definition of the derivative at a point, we have

$$f'(0) = \lim_{h\to 0} \frac{f(0+h)-f(0)}{h}. \tag{2.1.18}$$

Since $x = 0$ is the point where our function's definition switches from one piece to other, we need to consider left and right-hand limits. Consider the following, where we compute the left and right hand limits side by side.

$$\begin{array}{r|r} \lim\limits_{h\to 0^-} \dfrac{f(0+h)-f(0)}{h} = & \lim\limits_{h\to 0^+} \dfrac{f(0+h)-f(0)}{h} = \\ \lim\limits_{h\to 0^-} \dfrac{-h-0}{h} = & \lim\limits_{h\to 0^+} \dfrac{h-0}{h} = \\ \lim\limits_{h\to 0^-} -1 = -1 & \lim\limits_{h\to 0^+} 1 = 1 \end{array}$$

The last lines of each column tell the story: the left and right hand limits are not equal. Therefore the limit does not exist at 0, and f is not differentiable at 0. So we have

$$f'(x) = \begin{cases} -1 & x < 0 \\ 1 & x > 0 \end{cases}. \tag{2.1.19}$$

At $x = 0$, $f'(x)$ does not exist; there is a jump discontinuity at 0; see Figure 2.7. So $f(x) = |x|$ is differentiable everywhere except at 0.

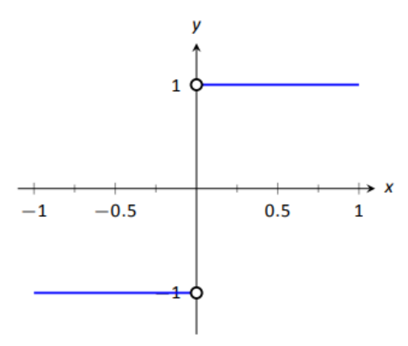


Figure 2.7: A graph of the derivative of $f(x) = |x|$.

The point of non-differentiability came where the piecewise defined function switched from one piece to the other. Our next example shows that this does not always cause trouble.

Example 40: Finding the derivative of a piecewise defined function

Find the derivative of $f(x)$, where $f(x) = \begin{cases} \sin x & x \le \pi/2 \\ 1 & x > \pi/2 \end{cases}$. See Figure 2.8.

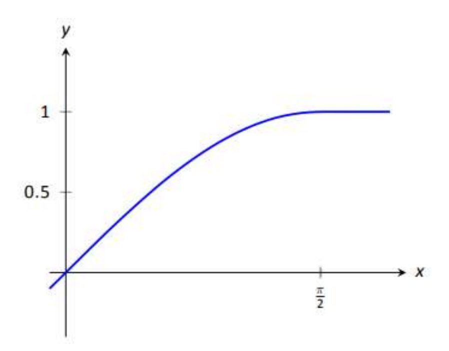


Figure 2.8: A graph of $f(x)$ as defined in Example 40.

Solution

Using Example 38, we know that when $x < \pi/2$, $f'(x) = \cos x$. It is easy to verify that when $x > \pi/2$, $f'(x) = 0$; consider:

$$\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}=\lim_{h\to 0}\frac{1-1}{h}=\lim_{h\to 0}0=0. \tag{2.1.20}$$

So far we have

$$f'(x)=\begin{cases}\cos x & x<\pi/2\\ 0 & x>\pi/2\end{cases}. \tag{2.1.21}$$

We still need to find $f'(\pi/2)$. Notice at $x=\pi/2$ that both pieces of f' are 0, meaning we can state that $f'(\pi/2)=0$.

Being more rigorous, we can again evaluate the difference quotient limit at $x=\pi/2$, utilizing again left and right--hand limits:

$$\begin{aligned}\lim_{h\to 0-}\frac{f(\pi/2+h)-f(\pi/2)}{h}&=\\ \lim_{h\to 0-}\frac{\sin(\pi/2+h)-\sin(\pi/2)}{h}&=\\ \lim_{h\to 0-}\frac{\sin(\frac{\pi}{2})\cos(h)+\sin(h)\cos(\frac{\pi}{2})-\sin(\frac{\pi}{2})}{h}&=\\ \lim_{h\to 0-}\frac{1\cdot\cos(h)+\sin(h)\cdot 0-1}{h}&=\\ &0\end{aligned}
\qquad\Bigg|\qquad
\begin{aligned}\lim_{h\to 0+}\frac{f(\pi/2+h)-f(\pi/2)}{h}&=\\ \lim_{h\to 0+}\frac{1-1}{h}&=\\ \lim_{h\to 0+}\frac{0}{h}&=\\ 0&\end{aligned}$$

Since both the left and right hand limits are 0 at $x=\pi/2$, the limit exists and $f'(\pi/2)$ exists (and is 0). Therefore we can fully write f' as

$$f'(x)=\begin{cases}\cos x & x\le\pi/2\\ 0 & x>\pi/2\end{cases}. \tag{2.1.22}$$

See Figure 2.9 for a graph of this function.

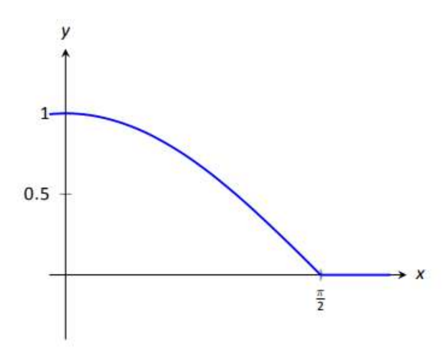


Figure 2.9: A graph of $f'(x)$ in Example 40.

Recall we pseudo--defined a continuous function as one in which we could sketch its graph without lifting our pencil. We can give a pseudo--definition for differentiability as well: it is a continuous function that does not have any "sharp corners." One such sharp corner is shown in Figure 2.6. Even though the function f in Example 40 is piecewise--defined, the transition is "smooth" hence it is differentiable. Note how in the graph of f in Figure 2.8 it is difficult to tell when f switches from one piece to the other; there is no "corner."

This section defined the derivative; in some sense, it answers the question of "What *is* the derivative?" The next section addresses the question "What does the derivative *mean*?"

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 2.1: Instantaneous Rates of Change- The Derivative is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

2.2: Interpretations of the Derivative

The previous section defined the derivative of a function and gave examples of how to compute it using its definition (i.e., using limits). The section also started with a brief motivation for this definition, that is, finding the instantaneous velocity of a falling object given its position function. The next section will give us more accessible tools for computing the derivative, tools that are easier to use than repeated use of limits.

This section falls in between the "What is the definition of the derivative?" and "How do I compute the derivative?" sections. Here we are concerned with "What does the derivative mean?", or perhaps, when read with the right emphasis, "What *is* the derivative?" We offer two interconnected interpretations of the derivative, hopefully explaining why we care about it and why it is worthy of study.

Interpretation of the Derivative #1: Instantaneous Rate of Change

The previous section started with an example of using the position of an object (in this case, a falling amusement--park rider) to find the object's velocity. This type of example is often used when introducing the derivative because we tend to readily recognize that velocity is the *instantaneous rate of change of position*. In general, if f is a function of x, then $f'(x)$ measures the instantaneous rate of change of f with respect to x. Put another way, the derivative answers "When x changes, at what rate does f change?" Thinking back to the amusement--park ride, we asked "When time changed, at what rate did the height change?" and found the answer to be "By -64 feet per second."

Now imagine driving a car and looking at the speedometer, which reads "60 mph." Five minutes later, you wonder how far you have traveled. Certainly, lots of things could have happened in those 5 minutes; you could have intentionally sped up significantly, you might have come to a complete stop, you might have slowed to 20 mph as you passed through construction. But suppose that you know, as the driver, none of these things happened. You know you maintained a fairly consistent speed over those 5 minutes. What is a good approximation of the distance traveled?

One could argue the *only* good approximation, given the information provided, would be based on "distance = rate $\times$ time." In this case, we assume a constant rate of 60 mph with a time of $5/60$ hours. Hence we would approximate the distance traveled as 5 miles.

Referring back to the falling amusement--park ride, knowing that at $t = 2$ the velocity was -64 ft/s, we could reasonably assume that 1 second later the riders' height would have dropped by about 64 feet. Knowing that the riders were *accelerating* as they fell would inform us that this is an *under--approximation*. If all we knew was that $f(2) = 86$ and $f'(2) = -64$, we'd know that we'd have to stop the riders quickly otherwise they would hit the ground!

Units of the Derivative

It is useful to recognize the *units* of the derivative function. If y is a function of x, i.e., $y = f(x)$ for some function f, and y is measured in feet and x in seconds, then the units of $y' = f'$ are "feet per second," commonly written as "ft/s." In general, if y is measured in units P and x is measured in units Q, then y' will be measured in units "P per Q", or "P/Q." Here we see the fraction--like behavior of the derivative in the notation:

$$\text{the units of } \frac{dy}{dx} \text{ are } \frac{\text{units of } y}{\text{units of } x}. \tag{2.2.1}$$

Example 41: The meaning of the derivative: World Population

Let $P(t)$ represent the world population t minutes after 12:00 a.m., January 1, 2012. It is fairly accurate to say that $P(0) = 7,028,734,178$ (www.prb.org). It is also fairly accurate to state that $P'(0) = 156$; that is, at midnight on January 1, 2012, the population of the world was growing by about 156 *people per minute* (note the units). Twenty days later (or, 28,800 minutes later) we could reasonably assume the population grew by about $28,800 \cdot 156 = 4,492,800$ people.

Example 42: The meaning of the derivative: Manufacturing

The term *widget* is an economic term for a generic unit of manufacturing output. Suppose a company produces widgets and knows that the market supports a price of $10 per widget. Let $P(n)$ give the profit, in dollars, earned by manufacturing and selling n widgets. The company likely cannot make a (positive) profit making just one widget; the start--up costs will likely exceed $10. Mathematically, we would write this as $P(1) < 0$.

What do $P(1000)=500$ and $P'(1000)=0.25$ mean? Approximate $P(1100)$.

Solution:

The equation $P(1000)=500$ means that selling 1,000 widgets returns a profit of \$500. We interpret $P'(1000)=0.25$ as meaning that the profit is increasing at rate of \$0.25 per widget (the units are "dollars per widget.") Since we have no other information to use, our best approximation for $P(1100)$ is:

$$P(1100)\approx P(1000)+P'(1000)\times 100=\$500+100\cdot 0.25=\$525. \tag{2.2.2}$$

We approximate that selling 1,100 widgets returns a profit of \$525.

The previous examples made use of an important approximation tool that we first used in our previous "driving a car at 60 mph" example at the beginning of this section. Five minutes after looking at the speedometer, our best approximation for distance traveled assumed the rate of change was constant. In Examples 41 and 42 we made similar approximations. We were given rate of change information which we used to approximate total change. Notationally, we would say that

$$f(c+h)\approx f(c)+f'(c)\cdot h. \tag{2.2.3}$$

This approximation is best when h is "small." "Small" is a relative term; when dealing with the world population, h = 22 days = 28,800 minutes is small in comparison to years. When manufacturing widgets, 100 widgets is small when one plans to manufacture thousands.

The Derivative and Motion

One of the most fundamental applications of the derivative is the study of motion. Let $s(t)$ be a position function, where t is time and $s(t)$ is distance. For instance, s could measure the height of a projectile or the distance an object has traveled.

Let's let $s(t)$ measure the distance traveled, in feet, of an object after t seconds of travel. Then $s'(t)$ has units "feet per second," and $s'(t)$ measures the *instantaneous rate of distance change* -- it measures **velocity**.

Now consider $v(t)$, a velocity function. That is, at time t, $v(t)$ gives the velocity of an object. The derivative of v, $v'(t)$, gives the *instantaneous rate of velocity change* -- **acceleration**. (We often think of acceleration in terms of cars: a car may "go from 0 to 60 in 4.8 seconds." This is an *average* acceleration, a measurement of how quickly the velocity changed.) If velocity is measured in feet per second, and time is measured in seconds, then the units of acceleration (i.e., the units of $v'(t)$) are "feet per second per second," or $((\text{ft}/\text{s})/\text{s})$. We often shorten this to "feet per second squared," or ft/s^2, but this tends to obscure the meaning of the units.

Perhaps the most well known acceleration is that of gravity. In this text, we use $g=32\text{ft}/\text{s}^2$ or $g=9.8\text{m}/\text{s}^2$. What do these numbers mean?

A constant acceleration of $32(\text{ft}/\text{s})/\text{s}$ means that the velocity changes by 32ft/s each second. For instance, let $v(t)$ measures the velocity of a ball thrown straight up into the air, where v has units ft/s and t is measured in seconds. The ball will have a positive velocity while traveling upwards and a negative velocity while falling down. The acceleration is thus $-32\text{ft}/\text{s}^2$. If $v(1)=20\text{ft/s}$, then when $t=2$, the velocity will have decreased by 32ft/s; that is, $v(2)=-12\text{ft/s}$. We can continue: $v(3)=-44\text{ft/s}$, and we can also figure that $v(0)=52\text{ft/s}$.

These ideas are so important we write them out as a Key Idea.

Key Idea 1: The Derivative and Motion

1. Let $s(t)$ be the position function of an object. Then $s'(t)$ is the velocity function of the object.
2. Let $v(t)$ be the velocity function of an object. Then $v'(t)$ is the acceleration function of the object.

We now consider the second interpretation of the derivative given in this section. This interpretation is not independent from the first by any means; many of the same concepts will be stressed, just from a slightly different perspective.

Interpretation of the Derivative #2: The Slope of the Tangent Line

Given a function $y=f(x)$, the difference quotient $\frac{f(c+h)-f(c)}{h}$ gives a change in y values divided by a change in x values; i.e., it is a measure of the "rise over run," or "slope," of the line that goes through two points on the graph of f: $\big(c,f(c)\big)$ and $\big(c+h,f(c+h)\big)$.

As h shrinks to 0, these two points come close together; in the limit we find $f'(c)$, the slope of a special line called the tangent line that intersects f only once near $x = c$.

Lines have a constant rate of change, their slope. Nonlinear functions do not have a constant rate of change, but we can measure their *instantaneous rate of change* at a given x value c by computing $f'(c)$. We can get an idea of how f is behaving by looking at the slopes of its tangent lines. We explore this idea in the following example.

Example 43: Understanding the derivative: the rate of change

Consider $f(x) = x^2$ as shown in Figure 2.10. It is clear that at $x = 3$ the function is growing faster than at $x = 1$, as it is steeper at $x = 3$. How much faster is it growing?

Solution:

We can answer this directly after the following section, where we learn to quickly compute derivatives. For now, we will answer graphically, by considering the slopes of the respective tangent lines.

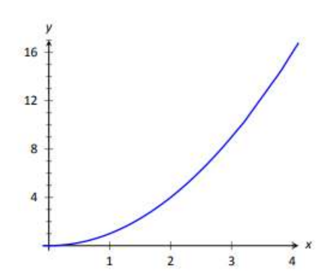


Figure 2.10: A graph of $f(x) = x^2$.

With practice, one can fairly effectively sketch tangent lines to a curve at a particular point. In Figure 2.11, we have sketched the tangent lines to f at $x = 1$ and $x = 3$, along with a grid to help us measure the slopes of these lines. At $x = 1$, the slope is 2; at $x = 3$, the slope is 6. Thus we can say not only is f growing faster at $x = 3$ than at $x = 1$, it is growing *three times as fast*.

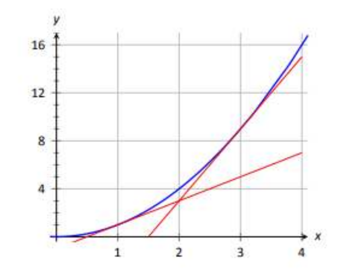


Figure 2.11: A graph of $f(x) = x^2$ and tangent lines.

Example 44: Understanding the graph of the derivative

Consider the graph of $f(x)$ and its derivative, $f'(x)$, in Figure 2.12(a). Use these graphs to find the slopes of the tangent lines to the graph of f at $x = 1$, $x = 2$, and $x = 3$.

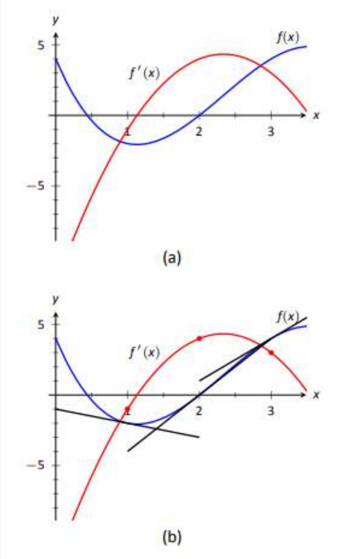


Figure 2.12: Graphs of f and f' in Example 44, along with tangent lines in (b).

Solution

To find the appropriate slopes of tangent lines to the graph of f, we need to look at the corresponding values of f'.

The slope of the tangent line to f at $x = 1$ is $f'(1)$; this looks to be about -1.
The slope of the tangent line to f at $x = 2$ is $f'(2)$; this looks to be about 4.
The slope of the tangent line to f at $x = 3$ is $f'(3)$; this looks to be about 3.
Using these slopes, the tangent lines to f are sketched in Figure 2.12(b). Included on the graph of f' in this figure are filled circles where $x = 1$, $x = 2$ and $x = 3$ to help better visualize the y value of f' at those points.

Example 45: Approximation with the derivative

Consider again the graph of $f(x)$ and its derivative $f'(x)$ in Example 44. Use the tangent line to f at $x = 3$ to approximate the value of $f(3.1)$.

Solution

Figure 2.13 shows the graph of f along with its tangent line, zoomed in at $x = 3$. Notice that near $x = 3$, the tangent line makes an excellent approximation of f. Since lines are easy to deal with, often it works well to approximate a function with its tangent line. (This is especially true when you don't actually know much about the function at hand, as we don't in this example.)

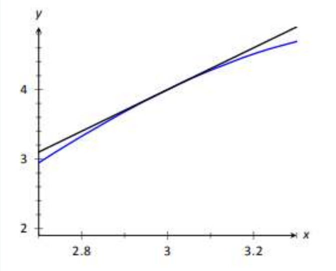


Figure 2.13: Zooming in on f at $x = 3$ for the function given in Examples 44 and 45.

While the tangent line to f was drawn in Example 44, it was not explicitly computed. Recall that the tangent line to f at $x = c$ is $y = f'(c)(x - c) + f(c)$. While f is not explicitly given, by the graph it looks like $f(3) = 4$. Recalling that $f'(3) = 3$, we can compute the tangent line to be approximately $y = 3(x - 3) + 4$. It is often useful to leave the tangent line in point--slope form.

To use the tangent line to approximate $f(3.1)$, we simply evaluate y at 3.1 instead of f.

$$f(3.1) \approx y(3.1) = 3(3.1 - 3) + 4 = .1 * 3 + 4 = 4.3. \tag{2.2.4}$$

We approximate $f(3.1) \approx 4.3$.

To demonstrate the accuracy of the tangent line approximation, we now state that in Example 45, $f(x) = -x^3 + 7x^2 - 12x + 4$. We can evaluate $f(3.1) = 4.279$. Had we known f all along, certainly we could have just made this computation. In reality, we often only know two things:

1. What $f(c)$ is, for some value of c, and
2. what $f'(c)$ is.

For instance, we can easily observe the location of an object and its instantaneous velocity at a particular point in time. We do not have a "function f" for the location, just an observation. This is enough to create an approximating function for f.

This last example has a direct connection to our approximation method explained above after Example 42. We stated that

$$f(c+h) \approx f(c) + f'(c) \cdot h. \tag{2.2.5}$$

If we know $f(c)$ and $f'(c)$ for some value $x = c$, then computing the tangent line at $(c, f(c))$ is easy: $y(x) = f'(c)(x-c) + f(c)$. In Example 45, we used the tangent line to approximate a value of f. Let's use the tangent line at $x = c$ to approximate a value of f near $x = c$; i.e., compute $y(c+h)$ to approximate $f(c+h)$, assuming again that h is "small." Note:

$$y(c+h) = f'(c)\big((c+h) - c\big) + f(c) = f'(c) \cdot h + f(c). \tag{2.2.6}$$

This is the exact same approximation method used above! Not only does it make intuitive sense, as explained above, it makes analytical sense, as this approximation method is simply using a tangent line to approximate a function's value.

The importance of understanding the derivative cannot be understated. When f is a function of x, $f'(x)$ measures the instantaneous rate of change of f with respect to x and gives the slope of the tangent line to f at x.

This page titled 2.2: Interpretations of the Derivative is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

2.3: Basic Differentiation Rules

The derivative is a powerful tool but is admittedly awkward given its reliance on limits. Fortunately, one thing mathematicians are good at is *abstraction*. For instance, instead of continually finding derivatives at a point, we abstracted and found the derivative function.

Let's practice abstraction on linear functions, $y = mx + b$. What is y'? Without limits, recognize that linear function are characterized by being functions with a constant rate of change (the slope). The derivative, y', gives the instantaneous rate of change; with a linear function, this is constant, m. Thus $y' = m$.

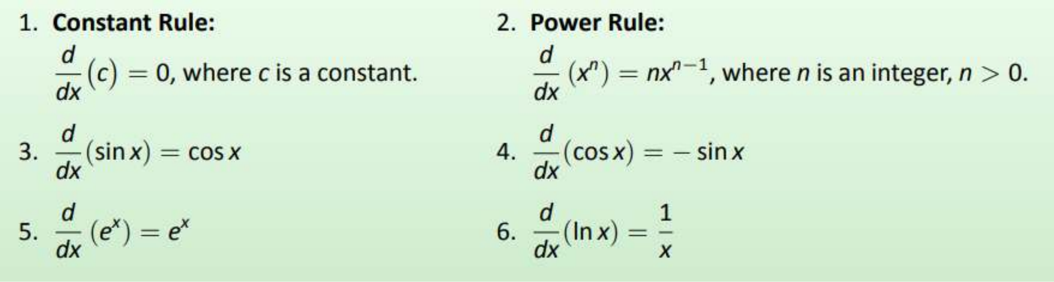
Let's abstract once more. Let's find the derivative of the general quadratic function, $f(x) = ax^2 + bx + c$. Using the definition of the derivative, we have:

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{a(x+h)^2 + b(x+h) + c - (ax^2+bx+c)}{h} \\ &= \lim_{h\to 0} \frac{ah^2 + 2ahx + bh}{h} \\ &= \lim_{h\to 0} ah + 2ax + b \\ &= 2ax + b. \end{aligned}$$

So if $y = 6x^2 + 11x - 13$, we can immediately compute $y' = 12x + 11$.

In this section (and in some sections to follow) we will learn some of what mathematicians have already discovered about the derivatives of certain functions and how derivatives interact with arithmetic operations. We start with a theorem.

Theorem 12: Derivatives of Common Functions



1. **Constant Rule:** $\frac{d}{dx}(c) = 0$, where c is a constant.
2. **Power Rule:** $\frac{d}{dx}(x^n) = nx^{n-1}$, where n is an integer, $n > 0$.
3. $\frac{d}{dx}(\sin x) = \cos x$
4. $\frac{d}{dx}(\cos x) = -\sin x$
5. $\frac{d}{dx}(e^x) = e^x$
6. $\frac{d}{dx}(\ln x) = \frac{1}{x}$

This theorem starts by stating an intuitive fact: constant functions have no rate of change as they are *constant*. Therefore their derivative is 0 (they change at the rate of 0). The theorem then states some fairly amazing things. The Power Rule states that the derivatives of Power Functions (of the form $y = x^n$) are very straightforward: multiply by the power, then subtract 1 from the power. We see something incredible about the function $y = e^x$: it is its own derivative. We also see a new connection between the sine and cosine functions.

One special case of the Power Rule is when $n = 1$, i.e., when $f(x) = x$. What is $f'(x)$? According to the Power Rule,

$$f'(x) = \frac{d}{dx}\big(x\big) = \frac{d}{dx}\big(x^1\big) = 1 \cdot x^0 = 1. \tag{2.3.1}$$

In words, we are asking "At what rate does f change with respect to x?" Since f *is* x, we are asking "At what rate does x change with respect to x?" The answer is: 1. They change at the same rate.

Let's practice using this theorem.

Example 46: Using Theorem 12 to find, and use, derivatives

Let $f(x) = x^3$.

1. Find $f'(x)$.
2. Find the equation of the line tangent to the graph of f at $x = -1$.
3. Use the tangent line to approximate $(-1.1)^3$.
4. Sketch f, f' and the found tangent line on the same axis.

Solution

1. The Power Rule states that if $f(x) = x^3$, then $f'(x) = 3x^2$.
2. To find the equation of the line tangent to the graph of f at $x = -1$, we need a point and the slope. The point is $(-1, f(-1)) = (-1, -1)$. The slope is $f'(-1) = 3$. Thus the tangent line has equation $y = 3(x-(-1)) + (-1) = 3x + 2$.
3. We can use the tangent line to approximate $(-1.1)^3$ as -1.1 is close to -1. We have

$$(-1.1)^3 \approx 3(-1.1) + 2 = -1.3. \tag{2.3.2}$$

   We can easily find the actual answer; $(-1.1)^3 = -1.331$.
4. See Figure 2.14.

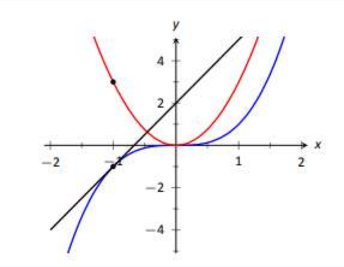


Figure 2.14: A graph of $f(x) = x^3$, along with its derivatives $f'(x) = 3x^2$ and its tangent line at $x = -1$.

Theorem 12 gives useful information, but we will need much more. For instance, using the theorem, we can easily find the derivative of $y = x^3$, but it does not tell how to compute the derivative of $y = 2x^3$, $y = x^3 + \sin x$ nor $y = x^3 \sin x$. The following theorem helps with the first two of these examples (the third is answered in the next section).

Theorem 13: Properties of the Derivative

Let f and g be differentiable on an open interval I and let c be a real number. Then:

1. **Sum/Difference Rule**:
   $\frac{d}{dx}\Big(f(x) \pm g(x)\Big) = \frac{d}{dx}\Big(f(x)\Big) \pm \frac{d}{dx}\Big(g(x)\Big) = f'(x) \pm g'(x)$
2. **Constant Multiple Rule:**
   $\frac{d}{dx}\Big(c \cdot f(x)\Big) = c \cdot \frac{d}{dx}\Big(f(x)\Big) = c \cdot f'(x)$.

Theorem 13 allows us to find the derivatives of a wide variety of functions. It can be used in conjunction with the Power Rule to find the derivatives of any polynomial. Recall in Example 36 that we found, using the limit definition, the derivative of $f(x) = 3x^2 + 5x - 7$. We can now find its derivative without expressly using limits:

$$\begin{aligned}\frac{d}{dx}\Big(3x^2 + 5x + 7\Big) &= 3\frac{d}{dx}\Big(x^2\Big) + 5\frac{d}{dx}\Big(x\Big) + \frac{d}{dx}\Big(7\Big)\\ &= 3 \cdot 2x + 5 \cdot 1 + 0\\ &= 6x + 5.\end{aligned}$$

We were a bit pedantic here, showing every step. Normally we would do all the arithmetic and steps in our head and readily find $\frac{d}{dx}\Big(3x^2 + 5x + 7\Big) = 6x + 5$.

Example 47: Using the tangent line to approximate a function value

Let $f(x) = \sin x + 2x + 1$. Approximate $f(3)$ using an appropriate tangent line.

Solution

This problem is intentionally ambiguous; we are to *approximate* using an *appropriate* tangent line. How good of an approximation are we seeking? What does appropriate mean?

In the "real world," people solving problems deal with these issues all time. One must make a judgment using whatever seems reasonable. In this example, the actual answer is $f(3) = \sin 3 + 7$, where the real problem spot is $\sin 3$. What is $\sin 3$?

Since 3 is close to π, we can assume $\sin 3 \approx \sin\pi = 0$. Thus one guess is $f(3) \approx 7$. Can we do better? Let's use a tangent line as instructed and examine the results; it seems best to find the tangent line at $x = \pi$.

Using Theorem 12 we find $f'(x) = \cos x + 2$. The slope of the tangent line is thus $f'(\pi) = \cos\pi + 2 = 1$. Also, $f(\pi) = 2\pi + 1 \approx 7.28$. So the tangent line to the graph of f at $x = \pi$ is $y = 1(x - \pi) + 2\pi + 1 = x + \pi + 1 \approx x + 4.14$. Evaluated at $x = 3$, our tangent line gives $y = 3 + 4.14 = 7.14$. Using the tangent line, our final approximation is that $f(3) \approx 7.14$.

Using a calculator, we get an answer accurate to 4 places after the decimal: $f(3) = 7.1411$. Our initial guess was 7; our tangent line approximation was more accurate, at 7.14.

The point is *not* "Here's a cool way to do some math without a calculator." Sure, that might be handy sometime, but your phone could probably give you the answer. Rather, the point is to say that tangent lines are a good way of approximating, and many scientists, engineers and mathematicians often face problems too hard to solve directly. So they approximate.

Higher Order Derivatives

The derivative of a function f is itself a function, therefore we can take its derivative. The following definition gives a name to this concept and introduces its notation.

Definition 11: Higher Order Derivatives

Let $y = f(x)$ be a differentiable function on I.

1. The *second derivative* of f is:

$$f''(x) = \frac{d}{dx}\Big(f'(x)\Big) = \frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{d^2y}{dx^2} = y''. \tag{2.3.3}$$

1. The *third derivative* of f is:

$$f'''(x) = \frac{d}{dx}\Big(f''(x)\Big) = \frac{d}{dx}\left(\frac{d^2y}{dx^2}\right) = \frac{d^3y}{dx^3} = y'''. \tag{2.3.4}$$

1. The n^{th} derivative} of f is:

$$f^{(n)}(x) = \frac{d}{dx}\left(f^{(n-1)}(x)\right) = \frac{d}{dx}\left(\frac{d^{n-1}y}{dx^{n-1}}\right) = \frac{d^ny}{dx^n} = y^{(n)}. \tag{2.3.5}$$

In general, when finding the fourth derivative and on, we resort to the $f^{(4)}(x)$ notation, not $f''''(x)$; after a while, too many ticks is too confusing.

Let's practice using this new concept.

Example 48: Finding higher order derivatives

Find the first four derivatives of the following functions:

1. $f(x) = 4x^2$
2. $f(x) = \sin x$
3. $f(x) = 5e^x$

Solution

1. Using the Power and Constant Multiple Rules, we have: $f'(x) = 8x$. Continuing on, we have

$$f''(x) = \frac{d}{dx}\Big(8x\Big) = 8; \qquad f'''(x) = 0; \qquad f^{(4)}(x) = 0. \tag{2.3.6}$$

   Notice how all successive derivatives will also be 0.
2. We employ Theorem 12 repeatedly.

$$f'(x) = \cos x; \qquad f''(x) = -\sin x; \qquad f'''(x) = -\cos x; \qquad f^{(4)}(x) = \sin x. \tag{2.3.7}$$

   Note how we have come right back to $f(x)$ again. (Can you quickly figure what $f^{(23)}(x)$ is?)
3. Employing Theorem 12 and the Constant Multiple Rule, we can see that

$$f'(x) = f''(x) = f'''(x) = f^{(4)}(x) = 5e^x. \tag{2.3.8}$$

Interpreting Higher Order Derivatives

What do higher order derivatives *mean*? What is the practical interpretation? Our first answer is a bit wordy, but is technically correct and beneficial to understand. That is,

$$\text{The second derivative of a function } f \text{ is the rate of change of the rate of change of } f. \tag{2.3.9}$$

One way to grasp this concept is to let f describe a position function. Then, as stated in Key Idea 1, f' describes the rate of position change: velocity. We now consider f'', which describes the rate of velocity change. Sports car enthusiasts talk of how fast a car can go from 0 to 60 mph; they are bragging about the *acceleration* of the car.

We started this chapter with amusement--park riders free--falling with position function $f(t) = -16t^2 + 150$. It is easy to compute $f'(t) = -32t$ ft/s and $f''(t) = -32$ (ft/s)/s. We may recognize this latter constant; it is the acceleration due to gravity. In keeping with the unit notation introduced in the previous section, we say the units are "feet per second per second." This is usually shortened to "feet per second squared," written as "ft/s^2."

It can be difficult to consider the meaning of the third, and higher order, derivatives. The third derivative is "the rate of change of the rate of change of the rate of change of f." That is essentially meaningless to the uninitiated. In the context of our position/velocity/acceleration example, the third derivative is the "rate of change of acceleration," commonly referred to as "jerk."

Make no mistake: higher order derivatives have great importance even if their practical interpretations are hard (or "impossible") to understand. The mathematical topic of *series* makes extensive use of higher order derivatives.

This page titled 2.3: Basic Differentiation Rules is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

2.4: The Product and Quotient Rules

The previous section showed that, in some ways, derivatives behave nicely. The Constant Multiple and Sum/Difference Rules established that the derivative of $f(x)=5x^2+\sin x$ was not complicated. We neglected computing the derivative of things like $g(x)=5x^2\sin x$ and $h(x)=\frac{5x^2}{\sin x}$ on purpose; their derivatives are *not* as straightforward. (If you had to guess what their respective derivatives are, you would probably guess wrong.) For these, we need the Product and Quotient Rules, respectively, which are defined in this section.

We begin with the Product Rule.

Theorem 14: Product Rule

Let f and g be differentiable functions on an open interval I. Then fg is a differentiable function on I, and

$$\frac{d}{dx}\Big(f(x)g(x)\Big)=f(x)g'(x)+f'(x)g(x). \tag{2.4.1}$$

Important: $\frac{d}{dx}\Big(f(x)g(x)\Big)\neq f'(x)g'(x)$! While this answer is simpler than the Product Rule, it is wrong.

We practice using this new rule in an example, followed by an example that demonstrates why this theorem is true.

Example 49: Using the Product Rule

Use the Product Rule to compute the derivative of $y=5x^2\sin x$. Evaluate the derivative at $x=\pi/2$.

Solution

To make our use of the Product Rule explicit, let's set $f(x)=5x^2$ and $g(x)=\sin x$. We easily compute/recall that $f'(x)=10x$ and $g'(x)=\cos x$. Employing the rule, we have

$$\frac{d}{dx}\Big(5x^2\sin x\Big)=5x^2\cos x+10x\sin x. \tag{2.4.2}$$

At $x=\pi/2$, we have

$$y'(\pi/2)=5\left(\frac{\pi}{2}\right)^2\cos\left(\frac{\pi}{2}\right)+10\frac{\pi}{2}\sin\left(\frac{\pi}{2}\right)=5\pi. \tag{2.4.3}$$

We graph y and its tangent line at $x=\pi/2$, which has a slope of 5π, in Figure 2.15. While this does not *prove* that the Product Rule is the correct way to handle derivatives of products, it helps validate its truth.

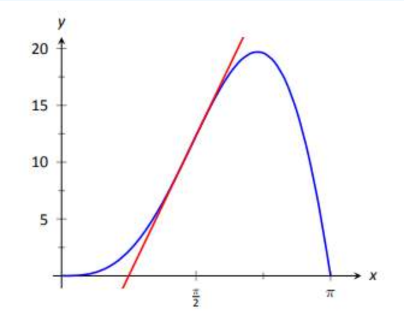


Figure 2.15: A graph of $y=5x^2\sin x$ and its tangent line at $x=\pi/2$.

We now investigate why the Product Rule is true.

Example 50: A proof of the Product Rule

Use the definition of the derivative to prove Theorem 14.

Solution

By the limit definition, we have

$$\frac{d}{dx}\Big(f(x)g(x)\Big)=\lim\frac{f(x+h)g(x+h)-f(x)g(x)}{}. \tag{2.4.4}$$

We now do something a bit unexpected; add 0 to the numerator (so that nothing is changed) in the form of $-f(x+h)g(x)+f(x+h)g(x)$, then do some regrouping as shown.

$$\begin{aligned}
\frac{d}{dx}\Big(f(x)g(x)\Big) &= \lim_{h\to 0}\frac{f(x+h)g(x+h)-f(x)g(x)}{h} \quad \text{(now add 0 to the numerator)}\\
&= \lim_{h\to 0}\frac{f(x+h)g(x+h)-f(x+h)g(x)+f(x+h)g(x)-f(x)g(x)}{h} \quad \text{(regroup)}\\
&= \lim_{h\to 0}\frac{\Big(f(x+h)g(x+h)-f(x+h)g(x)\Big)+\Big(f(x+h)g(x)-f(x)g(x)\Big)}{h}\\
&= \lim_{h\to 0}\frac{f(x+h)g(x+h)-f(x+h)g(x)}{h}\\
&\quad + \lim_{h\to 0}\frac{f(x+h)g(x)-f(x)g(x)}{h} \quad \text{(factor)}\\
&= \lim_{h\to 0} f(x+h)\frac{g(x+h)-g(x)}{h} + \lim_{h\to 0}\frac{f(x+h)-f(x)}{h}g(x) \quad \text{(apply limits)}\\
&= f(x)g'(x)+f'(x)g(x)
\end{aligned}$$

It is often true that we can recognize that a theorem is true through its proof yet somehow doubt its applicability to real problems. In the following example, we compute the derivative of a product of functions in two ways to verify that the Product Rule is indeed "right."

Example 51: Exploring alternate derivative methods

Let $y=(x^2+3x+1)(2x^2-3x+1)$. Find y^prime two ways: first, by expanding the given product and then taking the derivative, and second, by applying the Product Rule. Verify that both methods give the same answer.

Solution

We first expand the expression for y; a little algebra shows that $y=2x^4+3x^3-6x^2+1$. It is easy to compute y';

$$y' = 8x^3+9x^2-12x. \tag{2.4.5}$$

Now apply the Product Rule.

$$\begin{aligned}
y' &= (x^2+3x+1)(4x-3)+(2x+3)(2x^2-3x+1)\\
&= \Big(4x^3+9x^2-5x-3\Big)+\Big(4x^3-7x+3\Big)\\
&= 8x^3+9x^2-12x.
\end{aligned}$$

The uninformed usually assume that "the derivative of the product is the product of the derivatives." Thus we are tempted to say that $y'=(2x+3)(4x-3)=8x^2+6x-9$. Obviously this is not correct.

Example 52: Using the Product Rule with a product of these three functions

Let $y=x^3\ln x\cos x$. Find y'.

Solution

We have a product of three functions while the Product Rule only specifies how to handle a product of two functions. Our method of handling this problem is to simply group the latter two functions together, and consider $y=x^3\Big(\ln x\cos x\Big)$. Following the Product Rule, we have

$$\begin{aligned}
y' &= (x^3)\Big(\ln x\cos x\Big)' + 3x^2\Big(\ln x\cos x\Big)\\
&\text{To evaluate } \Big(\ln x\cos x\Big)'\text{, we apply the Product Rule again:}\\
&= (x^3)\Big(\ln x(-\sin x)+\frac{1}{x}\cos x\Big)+3x^2\Big(\ln x\cos x\Big)\\
&= x^3\ln x(-\sin x)+x^3\frac{1}{x}\cos x+3x^2\ln x\cos x
\end{aligned}$$

Recognize the pattern in our answer above: when applying the Product Rule to a product of three functions, there are three terms added together in the final derivative. Each terms contains only one derivative of one of the original functions, and each function's derivative shows up in only one term. It is straightforward to extend this pattern to finding the derivative of a product of 4 or more functions.

We consider one more example before discussing another derivative rule.

Example 53: Using the Product Rule

Find the derivatives of the following functions.

1. $f(x)=x\ln x$
2. $g(x)=x\ln x-x$.

Solution:

Recalling that the derivative of $\ln x$ is $1/x$, we use the Product Rule to find our answers.

1. $\frac{d}{dx}\Big(x\ln x\Big)=x\cdot 1/x+1\cdot\ln x=1+\ln x$.
2. Using the result from above, we compute

$$\frac{d}{dx}\Big(x\ln x-x\Big)=1+\ln x-1=\ln x. \tag{2.4.6}$$

This seems significant; if the natural log function $\ln x$ is an important function (it is), it seems worthwhile to know a function whose derivative is $\ln x$. We have found one. (We leave it to the reader to find another; a correct answer will be *very* similar to this one.)

We have learned how to compute the derivatives of sums, differences, and products of functions. We now learn how to find the derivative of a quotient of functions.

Theorem 15: Quotient Rule

Let f and g be functions defined on an open interval I, where $g(x)\neq 0$ on I. Then f/g is differentiable on I, and

$$\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right)=\frac{g(x)f'(x)-f(x)g'(x)}{g(x)^2}. \tag{2.4.7}$$

The Quotient Rule is not hard to use, although it might be a bit tricky to remember. A useful mnemonic works as follows. Consider a fraction's numerator and denominator as "HI" and "LO", respectively. Then

$$\frac{d}{dx}\left(\frac{\text{HI}}{\text{LO}}\right)=\frac{\text{LO}\cdot\text{dHI}-\text{HI}\cdot\text{dLO}}{\text{LOLO}}, \tag{2.4.8}$$

read "low dee high minus high dee low, over low low." Said fast, that phrase can roll off the tongue, making it easy to memorize. The "dee high" and "dee low" parts refer to the derivatives of the numerator and denominator, respectively.

Let's practice using the Quotient Rule.

Example 54: Using the Quotient Rule

Let $f(x)=\frac{5x^2}{\sin x}$. Find $f'(x)$.

Solution

Directly applying the Quotient Rule gives:

$$\begin{aligned}\frac{d}{dx}\left(\frac{5x^2}{\sin x}\right)&=\frac{\sin x\cdot 10x-5x^2\cdot\cos x}{\sin^2 x}\\&=\frac{10x\sin x-5x^2\cos x}{\sin^2 x}.\end{aligned}$$

The Quotient Rule allows us to fill in holes in our understanding of derivatives of the common trigonometric functions. We start with finding the derivative of the tangent function.

Example 55

Using the Quotient Rule to find $\frac{d}{dx}\left(\tan x\right)$.

Find the derivative of $y=\tan x$.

Solution

At first, one might feel unequipped to answer this question. But recall that $\tan x=\sin x/\cos x$, so we can apply the Quotient Rule.

$$\begin{aligned}\frac{d}{dx}\left(\tan x\right)&=\frac{d}{dx}\left(\frac{\sin x}{\cos x}\right)\\&=\frac{\cos x\cos x-\sin x(-\sin x)}{\cos^2 x}\\&=\frac{\cos^2 x+\sin^2 x}{\cos^2 x}\\&=\frac{1}{\cos^2 x}\\&=\sec^2 x.\end{aligned}$$

This is a beautiful result. To confirm its truth, we can find the equation of the tangent line to $y=\tan x$ at $x=\pi/4$. The slope is $\sec^2(\pi/4)=2$; $y=\tan x$, along with its tangent line, is graphed in Figure 2.16.

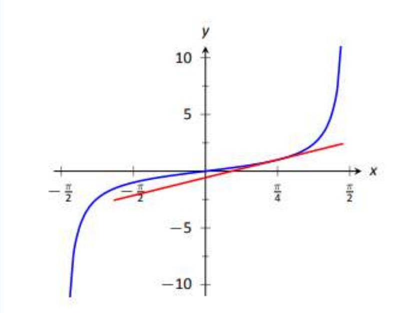


Figure 2.16: A graph of $y=\tan x$ along with its tangent line at $x=\pi/4$.

We include this result in the following theorem about the derivatives of the trigonometric functions. Recall we found the derivative of $y=\sin x$ in Example 38 and stated the derivative of the cosine function in Theorem 12. The derivatives of the cotangent, cosecant and secant functions can all be computed directly using Theorem 12 and the Quotient Rule.

Theorem 16: Derivatives of Trigonometric Functions

1. $\frac{d}{dx}\left(\sin x\right)=\cos x$
2. $\frac{d}{dx}\left(\cos x\right)=-\sin x$
3. $\frac{d}{dx}\left(\tan x\right)=\sec^2 x$
4. $\frac{d}{dx}\left(\cot x\right)=-\csc^2 x$
5. $\frac{d}{dx}\left(\sec x\right)=\sec x\tan x$
6. $\frac{d}{dx}\left(\csc x\right)=-\csc x\cot x$

To remember the above, it may be helpful to keep in mind that the derivatives of the trigonometric functions that start with "c" have a minus sign in them.

Example 56: Exploring alternative derivative methods

In Example 54 the derivative of $f(x)=\frac{5x^2}{\sin x}$ was found using the Quotient Rule. Rewriting f as $f(x)=5x^2\csc x$, find f' using Theorem 16 and verify the two answers are the same.}

Solution

We found in Example 54 that the $f'(x) = \frac{10x\sin x - 5x^2\cos x}{\sin^2 x}$. We now find f' using the Product Rule, considering f as $f(x) = 5x^2\csc x$.

$$\begin{aligned} f'(x) &= \frac{d}{dx}\left(5x^2\csc x\right) \\ &= 5x^2(-\csc x\cot x) + 10x\csc x \quad \text{(now rewrite trig functions)} \\ &= 5x^2\cdot\frac{-1}{\sin x}\cdot\frac{\cos x}{\sin x} + \frac{10x}{\sin x} \\ &= \frac{-5x^2\cos x}{\sin^2 x} + \frac{10x}{\sin x} \quad \text{(get common denominator)} \\ &= \frac{10x\sin x - 5x^2\cos x}{\sin^2 x} \end{aligned}$$

Finding f' using either method returned the same result. At first, the answers looked different, but some algebra verified they are the same. In general, there is not one final form that we seek; the immediate result from the Product Rule is fine. Work to "simplify" your results into a form that is most readable and useful to you.

The Quotient Rule gives other useful results, as show in the next example.

Example 57: Using the Quotient Rule to expand the Power Rule

Find the derivatives of the following functions.

1. $f(x) = \frac{1}{x}$
2. $f(x) = \frac{1}{x^n}$, where $n > 0$ is an integer.

Solution

We employ the Quotient Rule.

1. $f'(x) = \frac{x\cdot 0 - 1\cdot 1}{x^2} = -\frac{1}{x^2}$.
2. $f'(x) = \frac{x^n\cdot 0 - 1\cdot nx^{n-1}}{(x^n)^2} = -\frac{nx^{n-1}}{x^{2n}} = -\frac{n}{x^{n+1}}$.

The derivative of $y = \frac{1}{x^n}$ turned out to be rather nice. It gets better. Consider:

$$\begin{aligned} \frac{d}{dx}\left(\frac{1}{x^n}\right) &= \frac{d}{dx}\left(x^{-n}\right) \quad \text{(apply result from Example 57)} \\ &= -\frac{n}{x^{n+1}} \text{(rewrite algebraically)} \\ &= -nx^{-(n+1)} \\ &= -nx^{-n-1}. \end{aligned}$$

This is reminiscent of the Power Rule: multiply by the power, then subtract 1 from the power. We now add to our previous Power Rule, which had the restriction of $n > 0$.

Theorem 17: Power Rule with Integer Exponents

Let $f(x) = x^n$, where $n \neq 0$ is an integer. Then

$$f'(x) = n\cdot x^{n-1}. \tag{2.4.9}$$

Taking the derivative of many functions is relatively straightforward. It is clear (with practice) what rules apply and in what order they should be applied. Other functions present multiple paths; different rules may be applied depending on how the function is treated. One of the beautiful things about calculus is that there is not "the" right way; each path, when applied correctly, leads to the same result, the derivative. We demonstrate this concept in an example.

Example 58: Exploring alternate derivative methods

Let $f(x) = \frac{x^2 - 3x + 1}{x}$. Find $f'(x)$ in each of the following ways:

1. By applying the Quotient Rule,
2. by viewing f as $f(x) = \left(x^2 - 3x + 1\right)\cdot x^{-1}$ and applying the Product and Power Rules, and
3. by "simplifying\primeskip" first through division.

Verify that all three methods give the same result.

Solution

1. Applying the Quotient Rule gives:

$$f'(x)=\frac{x\cdot\left(2x-3\right)-\left(x^2-3x+1\right)\cdot 1}{x^2}=\frac{x^2-1}{x^2}=1-\frac{1}{x^2}. \tag{2.4.10}$$

2. By rewriting f, we can apply the Product and Power Rules as follows:

$$\begin{aligned}f'(x)&=\left(x^2-3x+1\right)\cdot(-1)x^{-2}+\left(2x-3\right)\cdot x^{-1}\\&=-\frac{x^2-3x+1}{x^2}+\frac{2x-3}{x}\\&=-\frac{x^2-3x+1}{x^2}+\frac{2x^2-3x}{x^2}\\&=\frac{x^2-1}{x^2}=1-\frac{1}{x^2},\end{aligned}$$

   the same result as above.

3. As $x\neq 0$, we can divide through by x first, giving $f(x)=x-3+\frac{1}{x}$. Now apply the Power Rule.

$$f'(x)=1-\frac{1}{x^2}, \tag{2.4.11}$$

   the same result as before.

Example 58 demonstrates three methods of finding f'. One is hard pressed to argue for a "best method" as all three gave the same result without too much difficulty, although it is clear that using the Product Rule required more steps. Ultimately, the important principle to take away from this is: reduce the answer to a form that seems "simple" and easy to interpret. In that example, we saw different expressions for f', including:

$$\begin{aligned}1-\frac{1}{x^2}&=\frac{x\cdot\left(2x-3\right)-\left(x^2-3x+1\right)\cdot 1}{x^2}\\&=\left(x^2-3x+1\right)\cdot(-1)x^{-2}+\left(2x-3\right)\cdot x^{-1}.\end{aligned} \tag{2.4.12}$$

They are equal; they are all correct; only the first is "clear." Work to make answers clear.

In the next section we continue to learn rules that allow us to more easily compute derivatives than using the limit definition directly. We have to memorize the derivatives of a certain set of functions, such as "the derivative of $\sin x$ is $\cos x$." The Sum/Difference, Constant Multiple, Power, Product and Quotient Rules show us how to find the derivatives of certain combinations of these functions. The next section shows how to find the derivatives when we *compose* these functions together.

This page titled 2.4: The Product and Quotient Rules is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

2.5: The Chain Rule

We have covered almost all of the derivative rules that deal with combinations of two (or more) functions. The operations of addition, subtraction, multiplication (including by a constant) and division led to the Sum and Difference rules, the Constant Multiple Rule, the Power Rule, the Product Rule and the Quotient Rule. To complete the list of differentiation rules, we look at the last way two (or more) functions can be combined: the process of composition (i.e. one function "inside" another).

One example of a composition of functions is $f(x)=\cos(x^2)$. We currently do not know how to compute this derivative. If forced to guess, one would likely guess $f'(x)=-\sin(2x)$,where we recognize $-\sin x$ as the derivative of $\cos x$ and $2x$ as the derivative of x^2. However, this is not the case; $f'(x)\neq -\sin(2x)$. In Example 62 we'll see the correct answer, which employs the new rule this section introduces, the **Chain Rule**.

Before we define this new rule, recall the notation for composition of functions. We write $(f\circ g)(x)$ or $f(g(x))$,read as "f of g of x," to denote composing f with g. In shorthand, we simply write $f\circ g$ or $f(g)$ and read it as "f of g." Before giving the corresponding differentiation rule, we note that the rule extends to multiple compositions like $f(g(h(x)))$ or $f(g(h(j(x))))$,etc.

To motivate the rule, let's look at three derivatives we can already compute.

Example 59: Exploring similar derivatives

Find the derivatives of

a. $F_1(x)=(1-x)^2$,
b. $F_2(x)=(1-x)^3$, and
c. $F_3(x)=(1-x)^4$.

We'll see later why we are using subscripts for different functions and an uppercase F.

Solution

In order to use the rules we already have, we must first expand each function as

a. $F_1(x)=1-2x+x^2$,
b. $F_2(x)=1-3x+3x^2-x^3$ and
c. $F_3(x)=1-4x+6x^2-4x^3+x^4$.

It is not hard to see that:

$$\begin{aligned}F_1'(x)&=-2+2x\\ F_2'(x)&=-3+6x-3x^2\\ F_3'(x)&=-4+12x-12x^2+4x^3.\end{aligned}$$

An interesting fact is that these can be rewritten as

$$F_1'(x)=-2(1-x),\quad F_2'(x)=-3(1-x)^2 \text{ and } F_3'(x)=-4(1-x)^3. \tag{2.5.1}$$

A pattern might jump out at you. Recognize that each of these functions is a composition, letting $g(x)=1-x$:

$$\begin{aligned}F_1(x)&=f_1(g(x)),\text{ where } f_1(x)=x^2,\\ F_2(x)&=f_2(g(x)),\text{ where } f_2(x)=x^3,\\ F_3(x)&=f_3(g(x)),\text{ where } f_3(x)=x^4.\end{aligned}$$

We'll come back to this example after giving the formal statements of the Chain Rule; for now, we are just illustrating a pattern.

Theorem 18: The Chain Rule

Let $y=f(u)$ be a differentiable function of u and let $u=g(x)$ be a differentiable function of x. Then $y=f(g(x))$ is a differentiable function of x,and

$$y'=f'(g(x))\cdot g'(x). \tag{2.5.2}$$

To help understand the Chain Rule, we return to Example 59.

Example 60: Using the Chain Rule

Use the Chain Rule to find the derivatives of the following functions, as given in Example 59.

Solution

Example 59 ended with the recognition that each of the given functions was actually a composition of functions. To avoid confusion, we ignore most of the subscripts here.

$F_1(x) = (1-x)^2$:

We found that

$$y = (1-x)^2 = f(g(x)), \text{ where } f(x) = x^2 \text{ and } g(x) = 1-x. \tag{2.5.3}$$

To find y', we apply the Chain Rule. We need $f'(x) = 2x$ and $g'(x) = -1$.

Part of the Chain Rule uses $f'(g(x))$. This means substitute $g(x)$ for x in the equation for $f'(x)$. That is, $f'(x) = 2(1-x)$. Finishing out the Chain Rule we have

$$y' = f'(g(x)) \cdot g'(x) = 2(1-x) \cdot (-1) = -2(1-x) = 2x - 2. \tag{2.5.4}$$

$F_2(x) = (1-x)^3$:

Let $y = (1-x)^3 = f(g(x))$,where $f(x) = x^3$ and $g(x) = (1-x)$. We have $f'(x) = 3x^2$,so $f'(g(x)) = 3(1-x)^2$. The Chain Rule then states

$$y' = f'(g(x)) \cdot g'(x) = 3(1-x)^2 \cdot (-1) = -3(1-x)^2. \tag{2.5.5}$$

$F_3(x) = (1-x)^4$:

Finally, when $y = (1-x)^4$,we have $f(x) = x^4$ and $g(x) = (1-x)$. Thus $f'(x) = 4x^3$ and $f'(g(x)) = 4(1-x)^3$. Thus

$$y' = f'(g(x)) \cdot g'(x) = 4(1-x)^3 \cdot (-1) = -4(1-x)^3. \tag{2.5.6}$$

Example 60 demonstrated a particular pattern: when $f(x) = x^n$,then $y' = n \cdot (g(x))^{n-1} \cdot g'(x)$. This is called the Generalized Power Rule.

Theorem 19: Generalized Power Rule

Let $g(x)$ be a differentiable function and let $n \neq 0$ be an integer. Then

$$\frac{d}{dx}\Big(g(x)^n\Big) = n \cdot \big(g(x)\big)^{n-1} \cdot g'(x). \tag{2.5.7}$$

This allows us to quickly find the derivative of functions like $y = (3x^2 - 5x + 7 + \sin x)^{20}$. While it may look intimidating, the Generalized Power Rule states that

$$y' = 20(3x^2 - 5x + 7 + \sin x)^{19} \cdot (6x - 5 + \cos x). \tag{2.5.8}$$

Treat the derivative--taking process step--by--step. In the example just given, first multiply by 20, then rewrite the inside of the parentheses, raising it all to the 19th power. Then think about the derivative of the expression inside the parentheses, and multiply by that.

We now consider more examples that employ the Chain Rule.

Example 61: Using the Chain Rule

Find the derivatives of the following functions:

a. $y = \sin 2x$
b. $y = \ln(4x^3 - 2x^2)$
c. $y = e^{-x^2}$

Solution

a. Consider $y = \sin 2x$. Recognize that this is a composition of functions, where $f(x) = \sin x$ and $g(x) = 2x$. Thus

$$y' = f'(g(x)) \cdot g'(x) = \cos(2x) \cdot 2 = 2\cos 2x. \tag{2.5.9}$$

b. Recognize that $y = \ln(4x^3 - 2x^2)$ is the composition of $f(x) = \ln x$ and $g(x) = 4x^3 - 2x^2$. Also, recall that

$$\frac{d}{dx}\Big(\ln x\Big) = \frac{1}{x}. \tag{2.5.10}$$

This leads us to:

$$y' = \frac{1}{4x^3 - 2x^2} \cdot (12x^2 - 4x) = \frac{12x^2 - 4x}{4x^3 - 2x^2} = \frac{4x(3x-1)}{2x(2x^2 - x)} = \frac{2(3x-1)}{2x^2 - x}. \tag{2.5.11}$$

c. Recognize that $y = e^{-x^2}$ is the composition of $f(x) = e^x$ and $g(x) = -x^2$. Remembering that $f'(x) = e^x$,we have

$$y' = e^{-x^2} \cdot (-2x) = (-2x)e^{-x^2}. \tag{2.5.12}$$

Example 62: Using the Chain Rule to find a tangent line

Let $f(x) = \cos x^2$. Find the equation of the line tangent to the graph of f at $x = 1$.

Solution

The tangent line goes through the point $(1, f(1)) \approx (1, 0.54)$ with slope $f'(1)$. To find f',we need the Chain Rule.

$f'(x) = -\sin(x^2) \cdot (2x) = -2x\sin x^2$. Evaluated at $x = 1$,we have $f'(1) = -2\sin 1 \approx -1.68$. Thus the equation of the tangent line is

$$y = -1.68(x-1) + 0.54. \tag{2.5.13}$$

The tangent line is sketched along with f in Figure 2.17.

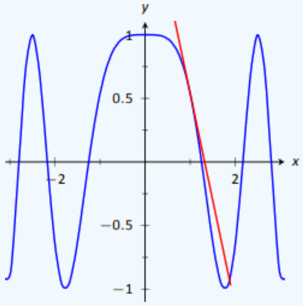


Figure 2.17: $f(x) = \cos x^2$ sketched along with its tangent line at $x = 1$.

The Chain Rule is used often in taking derivatives. Because of this, one can become familiar with the basic process and learn patterns that facilitate finding derivatives quickly. For instance,

$$\frac{d}{dx}\Big(\ln(\text{anything})\Big) = \frac{1}{\text{anything}} \cdot (\text{anything})' = \frac{(\text{anything})'}{\text{anything}}. \tag{2.5.14}$$

A concrete example of this is

$$\frac{d}{dx}\Big(\ln(3x^{15} - \cos x + e^x)\Big) = \frac{45x^{14} + \sin x + e^x}{3x^{15} - \cos x + e^x}. \tag{2.5.15}$$

While the derivative may look intimidating at first, look for the pattern. The denominator is the same as what was inside the natural log function; the numerator is simply its derivative.

This pattern recognition process can be applied to lots of functions. In general, instead of writing "anything", we use u as a generic function of x. We then say

$$\frac{d}{dx}\Big(\ln u\Big) = \frac{u'}{u}. \tag{2.5.16}$$

The following is a short list of how the Chain Rule can be quickly applied to familiar functions.

1. $\frac{d}{dx}\left(u^n\right) = n \cdot u^{n-1} \cdot u'.$
2. $\frac{d}{dx}\left(e^u\right) = u' \cdot e^u.$
3. $\frac{d}{dx}\left(\sin u\right) = u' \cdot \cos u.$
4. $\frac{d}{dx}\left(\cos u\right) = -u' \cdot \sin u.$
5. $\frac{d}{dx}\left(\tan u\right) = u' \cdot \sec^2 u.$

Of course, the Chain Rule can be applied in conjunction with any of the other rules we have already learned. We practice this next.

Example 63: Using the Product, Quotient and Chain Rules

Find the derivatives of the following functions.

a. $f(x) = x^5 \sin 2x^3$
b. $f(x) = \frac{5x^3}{e^{-x^2}}$.

Solution

a. We must use the Product and Chain Rules. Do not think that you must be able to "see" the whole answer immediately; rather, just proceed step--by--step.

$$f'(x) = x^5\left(6x^2 \cos 2x^3\right) + 5x^4\left(\sin 2x^3\right) = 6x^7 \cos 2x^3 + 5x^4 \sin 2x^3. \tag{2.5.17}$$

b. We must employ the Quotient Rule along with the Chain Rule. Again, proceed step--by--step.

$$\begin{aligned} f'(x) = \frac{e^{-x^2}\left(15x^2\right) - 5x^3\left((-2x)e^{-x^2}\right)}{\left(e^{-x^2}\right)^2} &= \frac{e^{-x^2}\left(10x^4 + 15x^2\right)}{e^{-2x^2}} \\ &= e^{x^2}\left(10x^4 + 15x^2\right). \end{aligned}$$

A key to correctly working these problems is to break the problem down into smaller, more manageable pieces. For instance, when using the Product and Chain Rules together, just consider the first part of the Product Rule at first: $f(x)g'(x)$. Just rewrite $f(x)$,then find $g'(x)$. Then move on to the $f'(x)g(x)$ part. Don't attempt to figure out both parts at once.

Likewise, using the Quotient Rule, approach the numerator in two steps and handle the denominator after completing that. Only simplify afterward.

We can also employ the Chain Rule itself several times, as shown in the next example.

Example 64: Using the Chain Rule multiple times

Find the derivative of $y = \tan^5(6x^3 - 7x)$.

Solution

Recognize that we have the $g(x) = \tan(6x^3 - 7x)$ function "inside" the $f(x) = x^5$ function; that is, we have $y = \left(\tan(6x^3 - 7x)\right)^5$. We begin using the Generalized Power Rule; in this first step, we do not fully compute the derivative. Rather, we are approaching this step--by--step.

$$y' = 5\left(\tan(6x^3 - 7x)\right)^4 \cdot g'(x). \tag{2.5.18}$$

We now find $g'(x)$. We again need the Chain Rule;

$$g'(x) = \sec^2(6x^3 - 7x) \cdot (18x^2 - 7). \tag{2.5.19}$$

Combine this with what we found above to give

$$\begin{aligned} y' &= 5\left(\tan(6x^3 - 7x)\right)^4 \cdot \sec^2(6x^3 - 7x) \cdot (18x^2 - 7) \\ &= (90x^2 - 35)\sec^2(6x^3 - 7x)\tan^4(6x^3 - 7x). \end{aligned}$$

This function is frankly a ridiculous function, possessing no real practical value. It is very difficult to graph, as the tangent function has many vertical asymptotes and $6x^3 - 7x$ grows so very fast. The important thing to learn from this is that the derivative can be found. In fact, it is not "hard;" one must take several simple steps and be careful to keep track of how to apply each of these steps.

It is a traditional mathematical exercise to find the derivatives of arbitrarily complicated functions just to demonstrate that it *can be done*. Just break everything down into smaller pieces.

Example 65: Using the Product, Quotient and Chain Rules

Find the derivative of $f(x)=\dfrac{x\cos(x^{-2})-\sin^2(e^{4x})}{\ln(x^2+5x^4)}$.

Solution

This function likely has no practical use outside of demonstrating derivative skills. The answer is given below without simplification. It employs the Quotient Rule, the Product Rule, and the Chain Rule three times.

$$f'(x) = \frac{\begin{array}{c}\Big(\ln(x^2+5x^4)\Big)\\ \cdot\Big[\Big(x\cdot(-\sin(x^{-2}))\cdot(-2x^{-3})+1\cdot\cos(x^{-2})\Big)-2\sin(e^{4x})\cdot\cos(e^{4x})\cdot(4e^{4x})\Big]\\ -\Big(x\cos(x^{-2})-\sin^2(e^{4x})\Big)\cdot\dfrac{2x+20x^3}{x^2+5x^4}\end{array}}{\Big(\ln(x^2+5x^4)\Big)^2}. \tag{2.5.20}$$

The reader is highly encouraged to look at each term and recognize why it is there. (I.e., the Quotient Rule is used; in the numerator, identify the "LOdHI" term, etc.) This example demonstrates that derivatives can be computed systematically, no matter how arbitrarily complicated the function is.

The Chain Rule also has theoretic value. That is, it can be used to find the derivatives of functions that we have not yet learned as we do in the following example.

Example 66: The Chain Rule and exponential functions

Use the Chain Rule to find the derivative of $y=a^x$ where $a>0,a\neq 1$ is constant.

Solution

We only know how to find the derivative of one exponential function: $y=e^x$; this problem is asking us to find the derivative of functions such as $y=2^x$.

This can be accomplished by rewriting a^x in terms of e. Recalling that e^x and $\ln x$ are inverse functions, we can write

$$a=e^{\ln a} \quad \text{and so} \quad y=a^x=e^{\ln(a^x)}.$$

By the exponent property of logarithms, we can "bring down" the power to get

$$y=a^x=e^{x(\ln a)}.$$

The function is now the composition $y=f(g(x))$,with $f(x)=e^x$ and $g(x)=x(\ln a)$. Since $f'(x)=e^x$ and $g'(x)=\ln a$, the Chain Rule gives

$$y'=e^{x(\ln a)}\cdot\ln a.$$

Recall that the $e^{x(\ln a)}$ term on the right hand side is just a^x,our original function. Thus, the derivative contains the original function itself. We have

$$y'=y\cdot\ln a=a^x\cdot\ln a.$$

The Chain Rule, coupled with the derivative rule of e^x,allows us to find the derivatives of all exponential functions.

The previous example produced a result worthy of its own "box."

Theorem 20: Derivatives of Exponential Functions

Let $f(x) = a^x$,for $a > 0, a \neq 1$. Then f is differentiable for all real numbers and

$$f'(x) = \ln a \cdot a^x.$$

Alternate Chain Rule Notation

It is instructive to understand what the Chain Rule "looks like" using "$\frac{dy}{dx}$" notation instead of y' notation. Suppose that $y = f(u)$ is a function of u,where $u = g(x)$ is a function of x,as stated in Theorem 18. Then, through the composition $f \circ g$,we can think of y as a function of x,as $y = f(g(x))$. Thus the derivative of y with respect to x makes sense; we can talk about $\frac{dy}{dx}$. This leads to an interesting progression of notation:

$$\begin{aligned} y' &= f'(g(x)) \cdot g'(x) \\ \frac{dy}{dx} &= y'(u) \cdot u'(x) \quad (\text{since } y = f(u) \text{ and } u = g(x)) \\ \frac{dy}{dx} &= \frac{dy}{du} \cdot \frac{du}{dx} \quad (\text{using "fractional" notation for the derivative}) \end{aligned}$$

Here the "fractional" aspect of the derivative notation stands out. On the right hand side, it seems as though the "du" terms cancel out, leaving

$$\frac{dy}{dx} = \frac{dy}{dx}. \tag{2.5.21}$$

It is important to realize that we *are not* canceling these terms; the derivative notation of $\frac{dy}{dx}$ is one symbol. It is equally important to realize that this notation was chosen precisely because of this behavior. It makes applying the Chain Rule easy with multiple variables. For instance,

$$\frac{dy}{dt} = \frac{dy}{d\bigcirc} \cdot \frac{d\bigcirc}{d\triangle} \cdot \frac{d\triangle}{dt}. \tag{2.5.22}$$

where $\bigcirc$ and $\triangle$ are any variables you'd like to use.

One of the most common ways of "visualizing" the Chain Rule is to consider a set of gears, as shown in Figure 2.18. The gears have 36, 18, and 6 teeth, respectively. That means for every revolution of the x gear, the u gear revolves twice. That is, the rate at which the u gear makes a revolution is twice as fast as the rate at which the x gear makes a revolution. Using the terminology of calculus, the rate of u-change, with respect to x,is $\frac{du}{dx} = 2$.

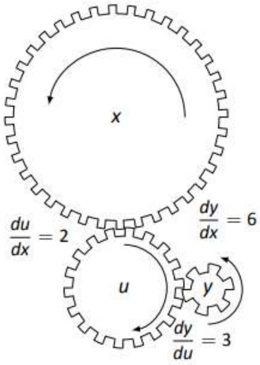


Figure 2.18: A series of gears to demonstrate the Chain Rule. Note how $\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$.

Likewise, every revolution of u causes 3 revolutions of y: $\frac{dy}{du} = 3$. How does y change with respect to x? For each revolution of x,y revolves 6 times; that is,

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx} = 2 \cdot 3 = 6. \tag{2.5.23}$$

We can then extend the Chain Rule with more variables by adding more gears to the picture.

It is difficult to overstate the importance of the Chain Rule. So often the functions that we deal with are compositions of two or more functions, requiring us to use this rule to compute derivatives. It is often used in practice when actual functions are unknown. Rather, through measurement, we can calculate $\frac{dy}{du}$ and $\frac{du}{dx}$. With our knowledge of the Chain Rule, finding $\frac{dy}{dx}$ is straightforward.

In the next section, we use the Chain Rule to justify another differentiation technique. There are many curves that we can draw in the plane that fail the "vertical line test." For instance, consider $x^2 + y^2 = 1$,which describes the unit circle. We may still be interested in finding slopes of tangent lines to the circle at various points. The next section shows how we can find $\frac{dy}{dx}$ without first "solving for y." While we can in this instance, in many other instances solving for y is impossible. In these situations, *implicit differentiation* is indispensable.

This page titled 2.5: The Chain Rule is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

2.6: Implicit Differentiation

In the previous sections we learned to find the derivative, $\frac{dy}{dx}$, or y', when y is given *explicitly* as a function of x. That is, if we know $y = f(x)$ for some function f, we can find y'. For example, given $y = 3x^2 - 7$, we can easily find $y' = 6x$. (Here we explicitly state how x and y are related. Knowing x, we can directly find y.)

Sometimes the relationship between y and x is not explicit; rather, it is *implicit*. For instance, we might know that $x^2 - y = 4$. This equality defines a relationship between x and y; if we know x, we could figure out y. Can we still find y'? In this case, sure; we solve for y to get $y = x^2 - 4$ (hence we now know y explicitly) and then differentiate to get $y' = 2x$.

Sometimes the *implicit* relationship between x and y is complicated. Suppose we are given $\sin(y) + y^3 = 6 - x^3$. A graph of this implicit function is given in Figure 2.19. In this case there is absolutely no way to solve for y in terms of elementary functions. The surprising thing is, however, that we can still find y' via a process known as **implicit differentiation**.

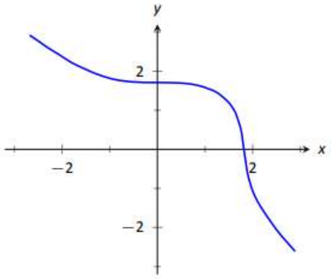


Figure 2.19: A graph of the implicit function $\sin(y) + y^3 = 6 - x^2$.

Implicit differentiation is a technique based on the Chain Rule that is used to find a derivative when the relationship between the variables is given implicitly rather than explicitly (solved for one variable in terms of the other).

We begin by reviewing the Chain Rule. Let f and g be functions of x. Then

$$\frac{d}{dx}\Big(f(g(x))\Big) = f'(g(x)) \cdot g'(x). \tag{2.6.1}$$

Suppose now that $y = g(x)$. We can rewrite the above as

$$\frac{d}{dx}\Big(f(y)\Big) = f'(y) \cdot y', \quad \text{or} \quad \frac{d}{dx}\Big(f(y)\Big) = f'(y) \cdot \frac{dy}{dx}. \tag{2.6.2}$$

These equations look strange; the key concept to learn here is that we can find y' even if we don't exactly know how y and x relate.

We demonstrate this process in the following example.

Example 67: Using Implicit Differentiation

Find y' given that $\sin(y) + y^3 = 6 - x^3$.

Solution

We start by taking the derivative of both sides (thus maintaining the equality.) We have :

$$\frac{d}{dx}\Big(\sin(y) + y^3\Big) = \frac{d}{dx}\Big(6 - x^3\Big). \tag{2.6.3}$$

The right hand side is easy; it returns $-3x^2$.

The left hand side requires more consideration. We take the derivative term--by--term. Using the technique derived from Equation 2.1 above, we can see that

$$\frac{d}{dx}\Big(\sin y\Big) = \cos y \cdot y'. \tag{2.6.4}$$

We apply the same process to the y^3 term.

$$\frac{d}{dx}\Big(y^3\Big) = \frac{d}{dx}\Big((y)^3\Big) = 3(y)^2 \cdot y'. \tag{2.6.5}$$

Putting this together with the right hand side, we have

$$\cos(y)y' + 3y^2y' = -3x^2. \tag{2.6.6}$$

Now solve for y'.

$$\begin{aligned} \cos(y)y' + 3y^2y' &= -3x^2. \\ \big(\cos y + 3y^2\big)y' &= -3x^2 \\ y' &= \frac{-3x^2}{\cos y + 3y^2} \end{aligned}$$

This equation for y' probably seems unusual for it contains both x and y terms. How is it to be used? We'll address that next.

Implicit functions are generally harder to deal with than explicit functions. With an explicit function, given an x value, we have an explicit formula for computing the corresponding y value. With an implicit function, one often has to find x and y values *at the same time* that satisfy the equation. It is much easier to demonstrate that a given point satisfies the equation than to actually find such a point.

For instance, we can affirm easily that the point $(\sqrt[3]{6}, 0)$ lies on the graph of the implicit function $\sin y + y^3 = 6 - x^3$. Plugging in 0 for y, we see the left hand side is 0. Setting $x = \sqrt[3]{6}$, we see the right hand side is also 0; the equation is satisfied. The following example finds the equation of the tangent line to this function at this point.

Example 68: Using Implicit Differentiation to find a tangent line

Find the equation of the line tangent to the curve of the implicitly defined function $\sin y + y^3 = 6 - x^3$ at the point $(\sqrt[3]{6}, 0)$.

Solution

In Example 67 we found that

$$y' = \frac{-3x^2}{\cos y + 3y^2}. \tag{2.6.7}$$

We find the slope of the tangent line at the point $(\sqrt[3]{6}, 0)$ by substituting $\sqrt[3]{6}$ for x and 0 for y. Thus at the point $(\sqrt[3]{6}, 0)$, we have the slope as

$$y' = \frac{-3(\sqrt[3]{6})^2}{\cos 0 + 3\cdot 0^2} = \frac{-3\sqrt[3]{36}}{1} \approx -9.91. \tag{2.6.8}$$

Therefore the equation of the tangent line to the implicitly defined function $\sin y + y^3 = 6 - x^3$ at the point $(\sqrt[3]{6}, 0)$ is

$$y = -3\sqrt[3]{36}(x - \sqrt[3]{6}) + 0 \approx -9.91x + 18. \tag{2.6.9}$$

The curve and this tangent line are shown in Figure 2.20.

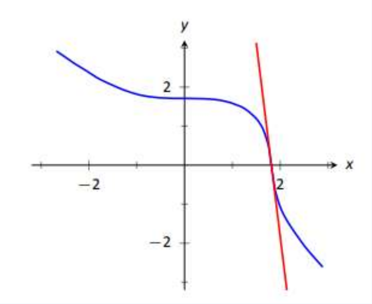


Figure 2.20: The function $\sin y + y^3 = 6 - x^2$ and its tangent line at the point $(\sqrt[3]{6}, 0)$.

This suggests a general method for implicit differentiation. For the steps below assume y is a function of x.

1. Take the derivative of each term in the equation. Treat the x terms like normal. When taking the derivatives of y terms, the usual rules apply except that, because of the Chain Rule, we need to multiply each term by y'.
2. Get all the y' terms on one side of the equal sign and put the remaining terms on the other side.
3. Factor out y'; solve for y' by dividing.

Practical Note: When working by hand, it may be beneficial to use the symbol $\frac{dy}{dx}$ instead of y', as the latter can be easily confused for y or y^1.

Example 69: Using Implicit Differentiation

Given the implicitly defined function $y^3 + x^2y^4 = 1 + 2x$, find y'.

Solution

We will take the implicit derivatives term by term. The derivative of y^3 is $3y^2y'$.

The second term, x^2y^4, is a little tricky. It requires the Product Rule as it is the product of two functions of x: x^2 and y^4. Its derivative is $x^2(4y^3y') + 2xy^4$. The first part of this expression requires a y' because we are taking the derivative of a y term. The second part does not require it because we are taking the derivative of x^2.

The derivative of the right hand side is easily found to be 2. In all, we get:

$$3y^2y' + 4x^2y^3y' + 2xy^4 = 2. \tag{2.6.10}$$

Move terms around so that the left side consists only of the y' terms and the right side consists of all the other terms:

$$3y^2y' + 4x^2y^3y' = 2 - 2xy^4. \tag{2.6.11}$$

Factor out y' from the left side and solve to get

$$y' = \frac{2 - 2xy^4}{3y^2 + 4x^2y^3}. \tag{2.6.12}$$

To confirm the validity of our work, let's find the equation of a tangent line to this function at a point. It is easy to confirm that the point $(0, 1)$ lies on the graph of this function. At this point, $y' = 2/3$. So the equation of the tangent line is $y = 2/3(x - 0) + 1$. The function and its tangent line are graphed in Figure 2.21.

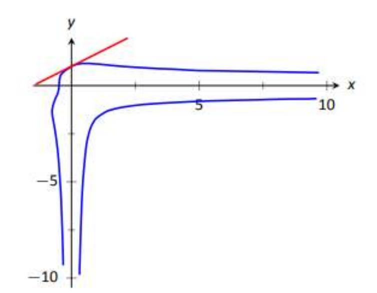


Figure 2.21: A graph of the implicitly defined function $y^3 + x^2y^4 = 1 + 2x$ along with its tangent line at the point $(0, 1)$.

Notice how our function looks much different than other functions we have seen. For one, it fails the vertical line test. Such functions are important in many areas of mathematics, so developing tools to deal with them is also important.

Example 70: Using Implicit Differentiation

Given the implicitly defined function $\sin(x^2y^2) + y^3 = x + y$, find y'.

Solution

Differentiating term by term, we find the most difficulty in the first term. It requires both the Chain and Product Rules.

$$\begin{aligned}\frac{d}{dx}\Big(\sin(x^2y^2)\Big) &= \cos(x^2y^2) \cdot \frac{d}{dx}\Big(x^2y^2\Big) \\ &= \cos(x^2y^2) \cdot \Big(x^2(2yy') + 2xy^2\Big) \\ &= 2(x^2yy' + xy^2)\cos(x^2y^2).\end{aligned}$$

We leave the derivatives of the other terms to the reader. After taking the derivatives of both sides, we have

$$2(x^2yy' + xy^2)\cos(x^2y^2) + 3y^2y' = 1 + y'. \tag{2.6.13}$$

We now have to be careful to properly solve for y', particularly because of the product on the left. It is best to multiply out the product. Doing this, we get

$$2x^2y\cos(x^2y^2)y' + 2xy^2\cos(x^2y^2) + 3y^2y' = 1 + y'. \tag{2.6.14}$$

From here we can safely move around terms to get the following:

$$2x^2y\cos(x^2y^2)y' + 3y^2y' - y' = 1 - 2xy^2\cos(x^2y^2). \tag{2.6.15}$$

Then we can solve for y' to get

$$y' = \frac{1 - 2xy^2\cos(x^2y^2)}{2x^2y\cos(x^2y^2) + 3y^2 - 1}. \tag{2.6.16}$$

A graph of this implicit function is given in Figure 2.22. It is easy to verify that the points $(0,0)$, $(0,1)$ and $(0,-1)$ all lie on the graph. We can find the slopes of the tangent lines at each of these points using our formula for y'.

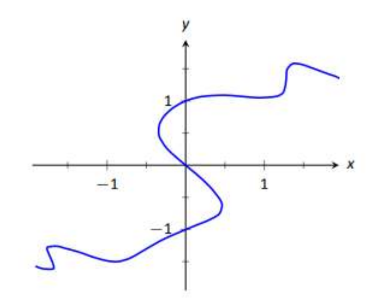


Figure 2.22: A graph of the implicitly defined function $\sin(x^2y^2) + y^3 = x + y$.

At $(0,0)$, the slope is -1.

At $(0,1)$, the slope is $1/2$.

At $(0,-1)$, the slope is also $1/2$.

The tangent lines have been added to the graph of the function in Figure 2.23.

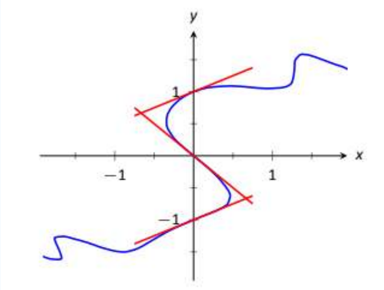


Figure 2.23: A graph of the implicitly defined function $\sin(x^2y^2) + y^3 = x + y$ and certain tangent lines.

Quite a few "famous" curves have equations that are given implicitly. We can use implicit differentiation to find the slope at various points on those curves. We investigate two such curves in the next examples.

Example 71: Finding slopes of tangent lines to a circle

Find the slope of the tangent line to the circle $x^2 + y^2 = 1$ at the point $(1/2, \sqrt{3}/2)$.

Solution

Taking derivatives, we get $2x + 2yy' = 0$. Solving for y' gives:

$$y' = \frac{-x}{y}. \tag{2.6.17}$$

This is a clever formula. Recall that the slope of the line through the origin and the point (x,y) on the circle will be y/x. We have found that the slope of the tangent line to the circle at that point is the opposite reciprocal of y/x, namely, $-x/y$. Hence these two lines are always perpendicular.

At the point $(1/2, \sqrt{3}/2)$, we have the tangent line's slope as

$$y' = \frac{-1/2}{\sqrt{3}/2} = \frac{-1}{\sqrt{3}} \approx -0.577. \tag{2.6.18}$$

A graph of the circle and its tangent line at $(1/2, \sqrt{3}/2)$ is given in Figure 2.24, along with a thin dashed line from the origin that is perpendicular to the tangent line. (It turns out that all normal lines to a circle pass through the center of the circle.)

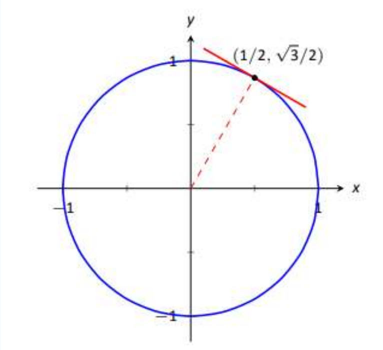


Figure 2.24: The unit circle with its tangent line at $(1/2, \sqrt{3}/2)$.

This section has shown how to find the derivatives of implicitly defined functions, whose graphs include a wide variety of interesting and unusual shapes. Implicit differentiation can also be used to further our understanding of "regular" differentiation.

One hole in our current understanding of derivatives is this: what is the derivative of the square root function? That is,

$$\frac{d}{dx}\left(\sqrt{x}\right) = \frac{d}{dx}\left(x^{1/2}\right) = ? \tag{2.6.19}$$

We allude to a possible solution, as we can write the square root function as a power function with a rational (or, fractional) power. We are then tempted to apply the Power Rule and obtain

$$\frac{d}{dx}\left(x^{1/2}\right) = \frac{1}{2}x^{-1/2} = \frac{1}{2\sqrt{x}}. \tag{2.6.20}$$

The trouble with this is that the Power Rule was initially defined only for positive integer powers, $n > 0$. While we did not justify this at the time, generally the Power Rule is proved using something called the Binomial Theorem, which deals only with positive integers. The Quotient Rule allowed us to extend the Power Rule to negative integer powers. Implicit Differentiation allows us to extend the Power Rule to rational powers, as shown below.

Let $y = x^{m/n}$, where m and n are integers with no common factors (so $m = 2$ and $n = 5$ is fine, but $m = 2$ and $n = 4$ is not). We can rewrite this explicit function implicitly as $y^n = x^m$. Now apply implicit differentiation.

$$\begin{aligned}
y &= x^{m/n} \\
y^n &= x^m \\
\frac{d}{dx}\left(y^n\right) &= \frac{d}{dx}\left(x^m\right) \\
n \cdot y^{n-1} \cdot y' &= m \cdot x^{m-1} \\
y' &= \frac{m}{n}\frac{x^{m-1}}{y^{n-1}} \quad \text{(now substitute } x^{m/n} \text{ for } y\text{)} \\
&= \frac{m}{n}\frac{x^{m-1}}{(x^{m/n})^{n-1}} \quad \text{(apply lots of algebra)} \\
&= \frac{m}{n}x^{(m-n)/n} \\
&= \frac{m}{n}x^{m/n-1}.
\end{aligned}$$

The above derivation is the key to the proof extending the Power Rule to rational powers. Using limits, we can extend this once more to include *all* powers, including irrational (even transcendental!) powers, giving the following theorem.

Theorem 21: Power Rule for Differentiation

Let $f(x) = x^n$, where $n \neq 0$ is a real number. Then f is a differentiable function, and $f'(x) = n \cdot x^{n-1}$.

This theorem allows us to say the derivative of x^π is $\pi x^{\pi-1}$.

We now apply this final version of the Power Rule in the next example, the second investigation of a "famous" curve.

Example 72: Using the Power Rule

Find the slope of $x^{2/3} + y^{2/3} = 8$ at the point $(8, 8)$.

Solution

This is a particularly interesting curve called an *astroid*. It is the shape traced out by a point on the edge of a circle that is rolling around inside of a larger circle, as shown in Figure 2.25.

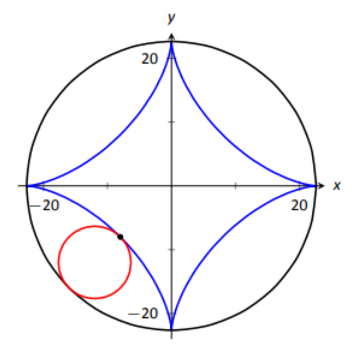


Figure 2.25: An astroid, traced out by a point on the smaller circle as it rolls inside the larger circle.

To find the slope of the astroid at the point $(8, 8)$, we take the derivative implicitly.

$$\begin{aligned} \frac{2}{3}x^{-1/3} + \frac{2}{3}y^{-1/3}y' &= 0 \\ \frac{2}{3}y^{-1/3}y' &= -\frac{2}{3}x^{-1/3} \\ y' &= -\frac{x^{-1/3}}{y^{-1/3}} \\ y' &= -\frac{y^{1/3}}{x^{1/3}} = -\sqrt[3]{\frac{y}{x}}. \end{aligned}$$

Plugging in $x = 8$ and $y = 8$, we get a slope of -1. The astroid, with its tangent line at $(8, 8)$, is shown in Figure 2.26.

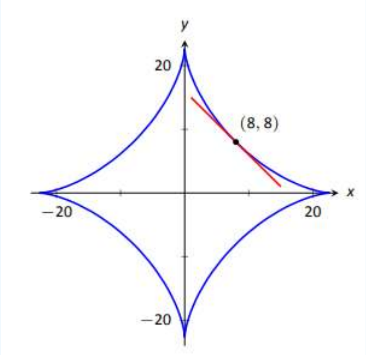


Figure 2.26: An astroid with a tangent line.

Implicit Differentiation and the Second Derivative

We can use implicit differentiation to find higher order derivatives. In theory, this is simple: first find $\frac{dy}{dx}$, then take its derivative with respect to x. In practice, it is not hard, but it often requires a bit of algebra. We demonstrate this in an example.

Example 73: Finding the second derivative

Given $x^2+y^2=1$, find $\frac{d^2y}{dx^2}=y''$.

Solution

We found that $y'=\frac{dy}{dx}=-x/y$ in Example 71. To find y'', we apply implicit differentiation to y'.

$$\begin{aligned} y'' &= \frac{d}{dx}\big(y'\big) \\ &= \frac{d}{dx}\left(-\frac{x}{y}\right) \qquad \text{(Now use the Quotient Rule.)} \\ &= -\frac{y(1)-x(y')}{y^2} \end{aligned}$$

replace y' with $-x/y$:

$$\begin{aligned} &= -\frac{y-x(-x/y)}{y^2} \\ &= -\frac{y+x^2/y}{y^2}. \end{aligned}$$

While this is not a particularly simple expression, it is usable. We can see that $y''>0$ when $y<0$ and $y''<0$ when $y>0$. In Section 3.4, we will see how this relates to the shape of the graph.

Logarithmic Differentiation

Consider the function $y=x^x$; it is graphed in Figure 2.27. It is well--defined for $x>0$ and we might be interested in finding equations of lines tangent and normal to its graph. How do we take its derivative?

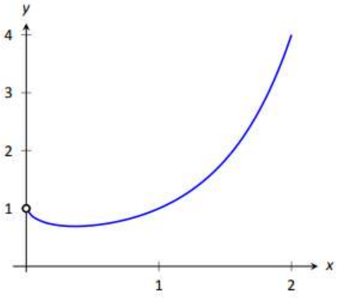


Figure 2.27: A plot of $y=x^x$.

The function is not a power function: it has a "power" of x, not a constant. It is not an exponential function: it has a "base" of x, not a constant.

A differentiation technique known as *logarithmic differentiation* becomes useful here. The basic principle is this: take the natural log of both sides of an equation $y=f(x)$, then use implicit differentiation to find y'. We demonstrate this in the following example.

Example 74: Using Logarithmic Differentiation

Given $y=x^x$, use logarithmic differentiation to find y'.

Solution

As suggested above, we start by taking the natural log of both sides then applying implicit differentiation.

$$\begin{aligned} y &= x^x \\ \ln(y) &= \ln(x^x) (\text{apply logarithm rule}) \\ \ln(y) &= x\ln x (\text{now use implicit differentiation}) \\ \frac{d}{dx}\Big(\ln(y)\Big) &= \frac{d}{dx}\Big(x\ln x\Big) \\ \frac{y'}{y} &= \ln x + x\cdot\frac{1}{x} \\ \frac{y'}{y} &= \ln x + 1 \\ y' &= y\Big(\ln x+1\Big)(\text{substitute } y=x^x) \\ y' &= x^x\Big(\ln x+1\Big). \end{aligned}$$

To "test" our answer, let's use it to find the equation of the tangent line at $x = 1.5$. The point on the graph our tangent line must pass through is $(1.5, 1.5^{1.5}) \approx (1.5, 1.837)$. Using the equation for y', we find the slope as

$$y' = 1.5^{1.5}\Big(\ln 1.5+1\Big) \approx 1.837(1.405) \approx 2.582. \tag{2.6.21}$$

Thus the equation of the tangent line is $y = 1.6833(x - 1.5) + 1.837$. Figure 2.28 graphs $y = x^x$ along with this tangent line.

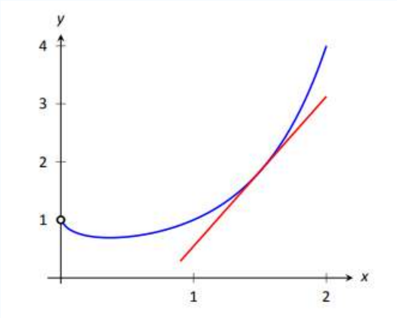


Figure 2.22: A graph of $y = x^x$ and its tangent line at $x = 1.5$.

Implicit differentiation proves to be useful as it allows us to find the instantaneous rates of change of a variety of functions. In particular, it extended the Power Rule to rational exponents, which we then extended to all real numbers. In the next section, implicit differentiation will be used to find the derivatives of *inverse* functions, such as $y = \sin^{-1} x$.

This page titled 2.6: Implicit Differentiation is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

2.7: Derivatives of Inverse Functions

Recall that a function $y = f(x)$ is said to be *one to one* if it passes the horizontal line test; that is, for two different x values x_1 and x_2, we do *not* have $f(x_1) = f(x_2)$. In some cases the domain of f must be restricted so that it is one to one. For instance, consider $f(x) = x^2$. Clearly, $f(-1) = f(1)$, so f is not one to one on its regular domain, but by restricting f to $(0, \infty)$, f is one to one.

Now recall that one to one functions have *inverses*. That is, if f is one to one, it has an inverse function, denoted by f^{-1}, such that if $f(a) = b$, then $f^{-1}(b) = a$. The domain of f^{-1} is the range of f, and vice-versa. For ease of notation, we set $g = f^{-1}$ and treat g as a function of x.

Since $f(a) = b$ implies $g(b) = a$, when we compose f and g we get a nice result:

$$f\big(g(b)\big) = f(a) = b. \tag{2.7.1}$$

In general, $f\big(g(x)\big) = x$ and $g\big(f(x)\big) = x$. This gives us a convenient way to check if two functions are inverses of each other: compose them and if the result is x, then they are inverses (on the appropriate domains.)

When the point (a, b) lies on the graph of f, the point (b, a) lies on the graph of g. This leads us to discover that the graph of g is the reflection of f across the line $y = x$. In Figure 2.29 we see a function graphed along with its inverse. See how the point $(1, 1.5)$ lies on one graph, whereas $(1.5, 1)$ lies on the other. Because of this relationship, whatever we know about f can quickly be transferred into knowledge about g.

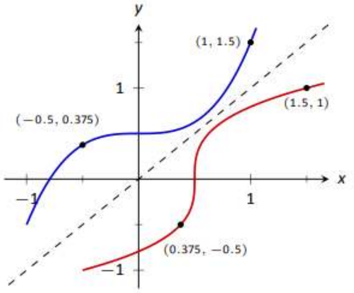


Figure 2.29: A function f along with its inverse f^{-1}. (Note how it does not matter which function we refer to as f*; the other is* f^{-1}*.)*

For example, consider Figure 2.30 where the tangent line to f at the point (a, b) is drawn. That line has slope $f'(a)$. Through reflection across $y = x$, we can see that the tangent line to g at the point (b, a) should have slope $\frac{1}{f'(a)}$. This then tells us that $g'(b) = \frac{1}{f'(a)}$.

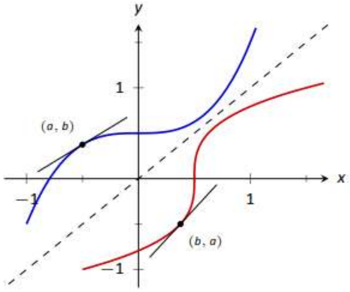
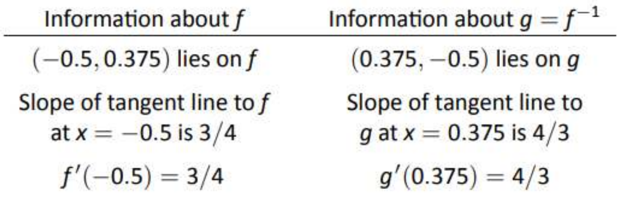


Figure 2.30: Corresponding tangent lines drawn to f and f^{-1}.

Consider:



Information about f	Information about $g = f^{-1}$
$(-0.5, 0.375)$ lies on f	$(0.375, -0.5)$ lies on g
Slope of tangent line to f at $x = -0.5$ is $3/4$	Slope of tangent line to g at $x = 0.375$ is $4/3$
$f'(-0.5) = 3/4$	$g'(0.375) = 4/3$

We have discovered a relationship between f' and g' in a mostly graphical way. We can realize this relationship analytically as well. Let $y=g(x)$, where again $g=f^{-1}$. We want to find y'. Since $y=g(x)$, we know that $f(y)=x$. Using the Chain Rule and Implicit Differentiation, take the derivative of both sides of this last equality.

$$\begin{aligned}\frac{d}{dx}\Big(f(y)\Big) &= \frac{d}{dx}\Big(x\Big)\\ f'(y)\cdot y' &= 1\\ y' &= \frac{1}{f'(y)}\\ y' &= \frac{1}{f'(g(x))}\end{aligned}$$

This leads us to the following theorem.

Theorem 22: Derivatives of Inverse Functions

Let f be differentiable and one to one on an open interval I, where $f'(x)\neq 0$ for all x in I, let J be the range of f on I, let g be the inverse function of f, and let $f(a)=b$ for some a in I. Then g is a differentiable function on J, and in particular,

1. $\left(f^{-1}\right)' (b) = g'(b) = \frac{1}{f'(a)}$\quad \text{ and }\quad 2. \left(f^{-1}\right)^\prime (x)=g^\prime (x) = \frac{1}{f^\prime(g(x))}\)

The results of Theorem 22 are not trivial; the notation may seem confusing at first. Careful consideration, along with examples, should earn understanding.

In the next example we apply Theorem 22 to the arcsine function.

Example 75: Finding the derivative of an inverse trigonometric function

Let $y=\arcsin x=\sin^{-1}x$. Find y' using Theorem 22.

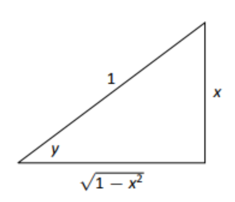


Figure 2.32: A right triangle defined by $y=\sin^{-1}(x/1)$ with the length of the third leg found using the Pythagorean Theorem.

Solution

Adopting our previously defined notation, let $g(x)=\arcsin x$ and $f(x)=\sin x$. Thus $f'(x)=\cos x$. Applying the theorem, we have

$$\begin{aligned}g'(x) &= \frac{1}{f'(g(x))}\\ &= \frac{1}{\cos(\arcsin x)}.\end{aligned}$$

This last expression is not immediately illuminating. Drawing a figure will help, as shown in Figure 2.32. Recall that the sine function can be viewed as taking in an angle and returning a ratio of sides of a right triangle, specifically, the ratio "opposite over hypotenuse." This means that the arcsine function takes as input a ratio of sides and returns an angle. The equation $y=\arcsin x$ can be rewritten as $y=\arcsin(x/1)$; that is, consider a right triangle where the hypotenuse has length 1 and the side opposite of the angle with measure y has length x. This means the final side has length $\sqrt{1-x^2}$, using the Pythagorean Theorem.

$$\text{Therefore }\cos(\sin^{-1}x)=\cos y=\sqrt{1-x^2}/1=\sqrt{1-x^2},\text{ resulting in} \tag{2.7.2}$$

$$\frac{d}{dx}\Big(\arcsin x\Big)=g'(x)=\frac{1}{\sqrt{1-x^2}}. \tag{2.7.3}$$

Remember that the input x of the arcsine function is a ratio of a side of a right triangle to its hypotenuse; the absolute value of this ratio will never be greater than 1. Therefore the inside of the square root will never be negative.

In order to make $y = \sin x$ one to one, we restrict its domain to $[-\pi/2, \pi/2]$; on this domain, the range is $[-1, 1]$. Therefore the domain of $y = \arcsin x$ is $[-1, 1]$ and the range is $[-\pi/2, \pi/2]$. When $x = \pm 1$, note how the derivative of the arcsine function is undefined; this corresponds to the fact that as $x \to \pm 1$, the tangent lines to arcsine approach vertical lines with undefined slopes.

In Figure 2.33 we see $f(x) = \sin x$ and $f^{-1} = \sin^{-1} x$ graphed on their respective domains. The line tangent to $\sin x$ at the point $(\pi/3, \sqrt{3}/2)$ has slope $\cos \pi/3 = 1/2$. The slope of the corresponding point on $\sin^{-1} x$, the point $(\sqrt{3}/2, \pi/3)$, is

$$\frac{1}{\sqrt{1-(\sqrt{3}/2)^2}} = \frac{1}{\sqrt{1-3/4}} = \frac{1}{\sqrt{1/4}} = \frac{1}{1/2} = 2, \tag{2.7.4}$$

verifying yet again that at corresponding points, a function and its inverse have reciprocal slopes.

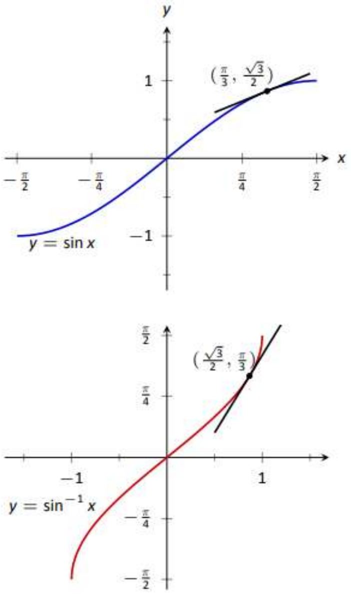


Figure 2.33: Graphs of $\sin x$ and $\sin^{-1} x$ along with corresponding tangent lines.

Using similar techniques, we can find the derivatives of all the inverse trigonometric functions. In Figure 2.31 we show the restrictions of the domains of the standard trigonometric functions that allow them to be invertible.

Function	Domain	Range	Inverse Function	Domain	Range
$\sin x$	$[-\pi/2, \pi/2]$	$[-1, 1]$	$\sin^{-1} x$	$[-1, 1]$	$[-\pi/2, \pi/2]$
$\cos x$	$[0, \pi]$	$[-1, 1]$	$\cos^{-1}(x)$	$[-1, 1]$	$[0, \pi]$
$\tan x$	$(-\pi/2, \pi/2)$	$(-\infty, \infty)$	$\tan^{-1}(x)$	$(-\infty, \infty)$	$(-\pi/2, \pi/2)$
$\csc x$	$[-\pi/2, 0) \cup (0, \pi/2]$	$(-\infty, -1] \cup [1, \infty)$	$\csc^{-1} x$	$(-\infty, -1] \cup [1, \infty)$	$[-\pi/2, 0) \cup (0, \pi/2]$
$\sec x$	$[0, \pi/2) \cup (\pi/2, \pi]$	$(-\infty, -1] \cup [1, \infty)$	$\sec^{-1}(x)$	$(-\infty, -1] \cup [1, \infty)$	$[0, \pi/2) \cup (\pi/2, \pi]$
$\cot x$	$(0, \pi)$	$(-\infty, \infty)$	$\cot^{-1}(x)$	$(-\infty, \infty)$	$(0, \pi)$

Figure 2.31: Domains and ranges of the trigonometric and inverse trigonometric functions.

Theorem 23: Derivatives of Inverse Trigonometric Functions

The inverse trigonometric functions are differentiable on all open sets contained in their domains (as listed in Figure 2.31) and their derivatives are as follows:

$$1. \frac{d}{dx}\left(\sin^{-1}(x)\right) = \frac{1}{\sqrt{1-x^2}} \qquad 4. \frac{d}{dx}\left(\cos^{-1}(x)\right) = -\frac{1}{\sqrt{1-x^2}} \tag{2.7.5}$$

$$2. \frac{d}{dx}\left(\sec^{-1}(x)\right) = \frac{1}{|x|\sqrt{x^2-1}} \qquad 5. \frac{d}{dx}\left(\csc^{-1}(x)\right) = -\frac{1}{|x|\sqrt{x^2-1}} \tag{2.7.6}$$

$$3.\frac{d}{dx}\left(\tan^{-1}(x)\right)=\frac{1}{1+x^2} \qquad 6.\frac{d}{dx}\left(\cot^{-1}(x)\right)=-\frac{1}{1+x^2} \tag{2.7.7}$$

Note how the last three derivatives are merely the opposites of the first three, respectively. Because of this, the first three are used almost exclusively throughout this text.

In Section 2.3, we stated without proof or explanation that $\frac{d}{dx}\left(\ln x\right)=\frac{1}{x}$. We can justify that now using Theorem 22, as shown in the example.

Example 76: Finding the derivative of y = ln x

Use Theorem 22 to compute $\frac{d}{dx}\left(\ln x\right)$.

Solution

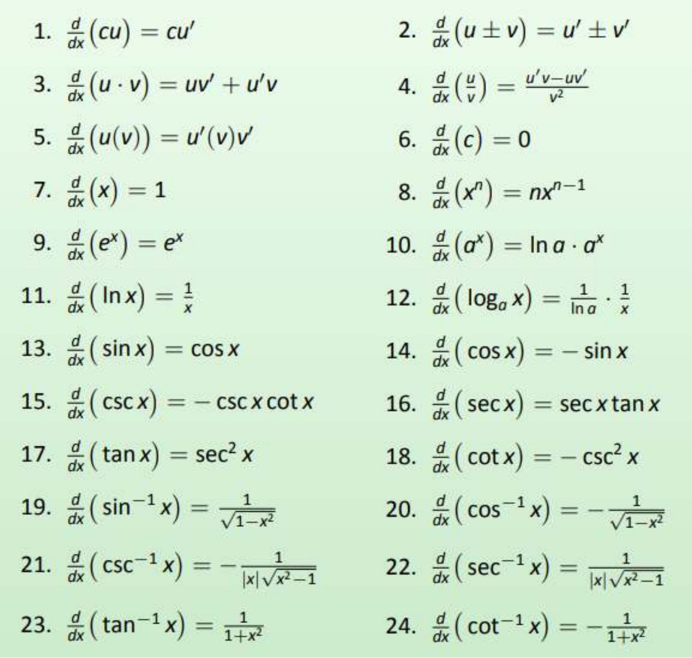
View $y=\ln x$ as the inverse of $y=e^x$. Therefore, using our standard notation, let $f(x)=e^x$ and $g(x)=\ln x$. We wish to find $g'(x)$. Theorem 22 gives:

$$\begin{aligned} g'(x) &= \frac{1}{f'(g(x))} \\ &= \frac{1}{e^{\ln x}} \\ &= \frac{1}{x}. \end{aligned}$$

In this chapter we have defined the derivative, given rules to facilitate its computation, and given the derivatives of a number of standard functions. We restate the most important of these in the following theorem, intended to be a reference for further work.

Theorem 24: Glossary of Derivatives of Elementary Functions

Let u and v be differentiable functions, and let a, c and n be real numbers, $a>0$, $n\neq 0$.



1. $\frac{d}{dx}(cu)=cu'$
2. $\frac{d}{dx}(u\pm v)=u'\pm v'$
3. $\frac{d}{dx}(u\cdot v)=uv'+u'v$
4. $\frac{d}{dx}\left(\frac{u}{v}\right)=\frac{u'v-uv'}{v^2}$
5. $\frac{d}{dx}(u(v))=u'(v)v'$
6. $\frac{d}{dx}(c)=0$
7. $\frac{d}{dx}(x)=1$
8. $\frac{d}{dx}(x^n)=nx^{n-1}$
9. $\frac{d}{dx}(e^x)=e^x$
10. $\frac{d}{dx}(a^x)=\ln a\cdot a^x$
11. $\frac{d}{dx}(\ln x)=\frac{1}{x}$
12. $\frac{d}{dx}(\log_a x)=\frac{1}{\ln a}\cdot\frac{1}{x}$
13. $\frac{d}{dx}(\sin x)=\cos x$
14. $\frac{d}{dx}(\cos x)=-\sin x$
15. $\frac{d}{dx}(\csc x)=-\csc x\cot x$
16. $\frac{d}{dx}(\sec x)=\sec x\tan x$
17. $\frac{d}{dx}(\tan x)=\sec^2 x$
18. $\frac{d}{dx}(\cot x)=-\csc^2 x$
19. $\frac{d}{dx}(\sin^{-1}x)=\frac{1}{\sqrt{1-x^2}}$
20. $\frac{d}{dx}(\cos^{-1}x)=-\frac{1}{\sqrt{1-x^2}}$
21. $\frac{d}{dx}(\csc^{-1}x)=-\frac{1}{|x|\sqrt{x^2-1}}$
22. $\frac{d}{dx}(\sec^{-1}x)=\frac{1}{|x|\sqrt{x^2-1}}$
23. $\frac{d}{dx}(\tan^{-1}x)=\frac{1}{1+x^2}$
24. $\frac{d}{dx}(\cot^{-1}x)=-\frac{1}{1+x^2}$

This page titled 2.7: Derivatives of Inverse Functions is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

2.E: Applications of Derivatives(Exercises)

2.1: Instantaneous Rates of Change: The Derivative

Terms and Concepts

1. T/F: Let f be a position function. The average rate of change on [a,b] is the slope of the line through the points $(a, f(a))$ and $(b, f(b))$.

2. T/F: The definition of the derivative of a function at a point involves taking a limit

3. In your own words, explain the difference between the average rate of change and instantaneous rate of change.

4. In your own words, explain the difference between Definitions 7 and 10.

5. Let $y = f(x)$. Give three different notations equivalent to "$f'(x)$."

Problems

In Exercises 6-12, use the definition of the derivative to compute the derivative of the given function.

6. $f(x)=6$

7. $f(x)=2x$

8. $f(t)=4-3t$

9. $g(x)=x^2$

10. $f(x)=3x^2-x+4$

11. $r(x)=\frac{1}{x}$

12. $r(s)=\frac{1}{s-2}$

In Exercises 13-19, a function and an x-value c are given.
(Note: these functions are the same as those given in Exercises 6 through 12.)

(a) Find the tangent line to the graph of the function at c.
(b) Find the normal line to the graph of the function at c.

13. $f(x)=6, \text{ at } x=-2.$

14. $f(x)=2x, \text{ at } x=3.$

15. $f(x)=4-3x, \text{ at } x=7.$

16. $g(x)=x^2, \text{ at } x=2.$

17. $f(x)=3x^2-x+4, \text{ at } x=-1.$

18. $r(x)=\frac{1}{x}, \text{ at } x=-2.$

19. $r(x)=\frac{1}{x-2}, \text{ at } x=3.$

In Exercises 20-23, a function f and an x-value a are given. Approximate the equation of the tangent line to the graph of f at $x=a$ by numerically approximating $f'(a)$, using $h=0.1$.

20. $f(x)=x^2+2x+1, \ x=3$

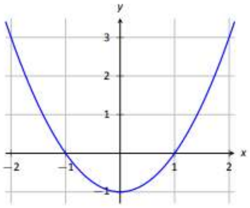
21. $f(x)=\frac{10}{x+1}, \ x=9$

22. $f(x)=e^x, \ x=2$

23. $f(x)=\cos x, \ x=0$

24. The graph of $f(x)=x^2-1$ is shown.
(a) Use the graph to approximate the slope of the tangent line to f at the following points: (-1,0),(0,-1) and (2,3).
(b) Using the definition, find $f'(x)$.

(c) Find the slope of the tangent line at the points (-1,0), (0,-1) and (2,3).

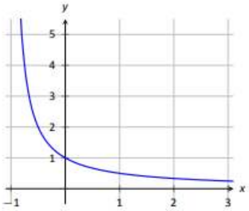


25. The graph of $f(x) = \frac{1}{x+1}$ is shown.

(a) Use the graph to approximate the slope of the tangent line to f at the following points: (0,1) and (1, 0.5).

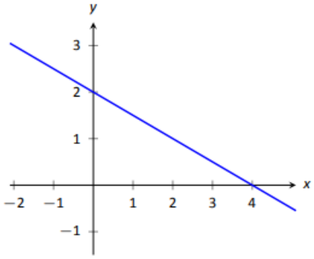
(b) Using the definition, find $f'(x)$.

(c) Find the slope of the tangent line at the points (0, 1) and (1, 0.5).

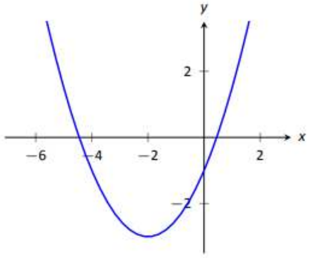


In Exercises 26-29, a graph of a function $f(x)$ is given. Using the graph, sketch $f'(x)$.

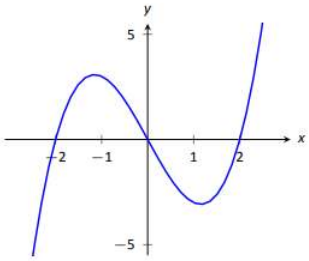
26.



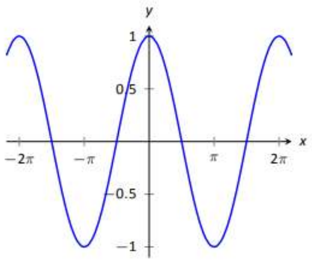
27.



28.



29.



30. Using the graph of $g(x)$ below, answer the following questions.
(a) Where is $g(x) > 0$?
(b) Where is $g(x) < 0$?
(c) Where is $g(x) = 0$?
(d) Where is $g'(x) < 0$?
(e) Where is $g'(x) > 0$?
(f) Where is $g'(x) = 0$?

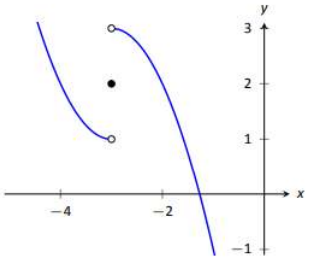
Review

31. Approximate $\lim\limits_{x\to 5} \frac{x^2+2x-35}{x^2-10.5+27.5}$.

32. Use the Bisection Method to approximate, accurate to two decimal places, the root of $g(x) = x^3 + x^2 + x - 1$ on [0.5, 0.6].

33. Give intervals on which each of the following functions are continuous.
(a) $\frac{1}{e^x+1}$
(b) $\frac{1}{e^x-1}$
(c) $\sqrt{5-x}$
(d) $\sqrt{5-x^2}$

34. Use the graph of $f(x)$ provided to answer the following.
(a) $\lim\limits_{x\to -3^-} f(x) =?$
(b) $\lim\limits_{x\to -3^+} f(x) =?$
(c) $\lim\limits_{x\to -3} f(x) =?$
(d) Where is f continuous?



2.2: Interpretations of the Derivative

Terms and Concepts

1. What is the instantaneous rate of change of position called?

2. Given a function $y = f(x)$, in your own words describe how to find the units of $f'(x)$.

3. What functions have a constant rate of change?

Problems

4. Given $f(5 = 10 \text{ and } f'(5) = 2$, approximate $f(6)$.

5. Given $P(100) = -67$ and $P'(100) = 5$, approximate $P(110)$.

6. Given $z(25) = 187$ and $z'(25) = 17$, approximate $z(20)$.

7. Knowing $f(10) = 25$ and $f'(10) = 5$ and the methods described in this section, which approximation is likely to be most accurate: $f(10.1),\ f(11),\ \text{or } f(20)$? Explain your reasoning.

8. Given $f(7) = 26 \text{ and } f(8) = 22$, approximate $f'(7)$.

9. Given $H(0) = 17$ and $H(2) = 29$, approximate $H'(2)$.

10. Let $V(x)$ measure the volume, in decibels, measured inside a restaurant with x customers. What are the units of $V'(x)$?

11. Let $v(t)$ measure the velocity, in ft/s, of a car moving in a straight line t seconds after starting. What are the units of $v'(t)$?

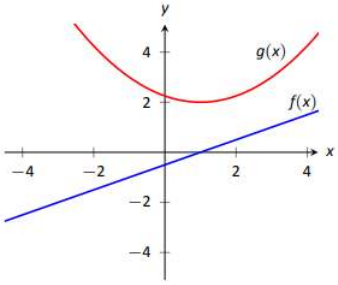
12. The height H, in feet, of a river is recorded t hours after midnight, April 1. What are the units of $H'(t)$?

13. P is the profit, in thousands of dollars, of producing and selling c cars.
(a) What are the units of $P'(c)$?
(b) What is likely true of $P(0P$?

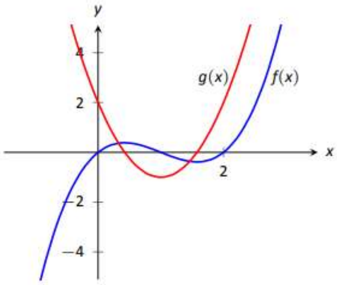
14. T is the temperature in degrees Fahrenheit, h hours after midnight on July 4 in Sidney, NE.
(a) What are the units of $T'(h)$?
(b) Is $T'(8)$ likely greater than or less than 0? Why?
(c) Is $T(8)$ likely greater than or less than 0? Why?

In Exercises 15-18, graphs of the functions $f(x) \text{ and } g(x)$ **are given. Identify which function is the derivative of the other.**

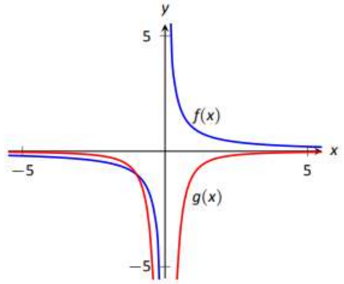
15.



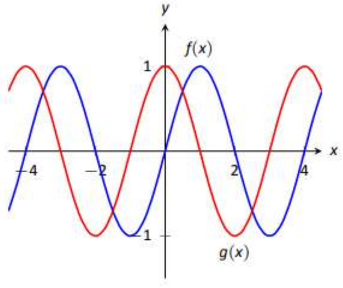
16.



17.



18.



Review

In Exercises 19-20, use the definition to compute the derivatives of the following functions.

19. $f(x) = 5x^2$

20. $f(x) = \sqrt{x}$ at $x = 9$.

In Exercises 21-22, numerically approximate the value of $f'(x)$ at the indicated x value.

21. $f(x) = \cos x$ at $x = \pi$.

22. $f(x) = \sqrt{x}$ at $x = 9$.

2.3: Basic Differentiation Rules

Terms and Concepts

1. What is the name of the rule which states that $\frac{d}{dx}(x^n) = nx^{n-1}$, where $n > 0$ is an integer?

2. What is $\frac{d}{dx}(\ln x)$?

3. Give an example of a function f(x) where $f'(x) = f(x)$.

4. Give an example of a function $f(x)$ where $f'(x) = 0$.

5. The derivative rules introduced in this section explain how to compute the derivative of which of the following functions?

- $f(x) = \frac{3}{x^2}$
- $g(x) = 3x^2 - x + 17$
- $h(x) = 5 \ln x$
- $j(x) = \sin x \cos x$
- $k(x) = e^{x^2}$
- $m(x) = \sqrt{x}$

6. Explain in your own words how to find the third derivative of a function $f(x)$.

7. Give an example of a function where $f'(x) \neq 0$ and $f''(x) = 0$.

8. Explain in your own words what the second derivative “means.”

9. If $f(x)$ describes a position function, then $f'(x)$ describes what kind of function? What kind of function is $f''(x)$?

10. Let $f(x)$ be a function measured in pounds, where x is measured in feet. What are the units of $f''(x)$?

Problems

In Exercises 11-25, compute the derivative of the given function.

11. $f(x) = 7x^2 - 5x + 7$

12. $g(x) = 14x^3 + 7x^2 + 11x - 29$

13. $m(t) = 9t^2 - \frac{1}{8}t^3 + 3t - 8$

14. $f(\theta) = 9\sin\theta + 10\cos\theta$

15. $f(r) = 6e^r$

16. $g(t) = 10t^4 - \cos t + 7\sin t$

17. $f(x) = 2\ln x - x$

18. $p(s) = \frac{1}{4}s^4 + \frac{1}{3}s^3 + \frac{1}{2}s^2 + s + 1$

19. $h(t) = e^t - \sin t - \cos t$

20. $f(x) = \ln(5x^2)$

21. $f(t) = \ln(17) + e^2 + \sin\pi/2$

22. $g(t) = (1 + 3t)^2$

23. $g(x) = (2x - 5)^3$

24. $f(x) = (1 - x)^3$

25. $f(x) = (2 - 3x)^2$

26. A property of logarithms is that $\log_a x = \frac{\log_b x}{\log_b a}$, for all bases a, b>0, $\neq 1$.
(a) Rewrite this identity when $b = e$, i.e., using $\log ex = \ln x$.
(b) Use part (a) to find the derivative of $y = \log_a x$.
(c) Give the derivative of $y = \log_{10} x$.

In Exercises 27-32, compute the first four derivatives of the given function.

27. $f(x) = x^6$

28. $g(x) = 2\cos x$

29. $h(t) = t^2 - e^t$

30. $p(\theta) = \theta^4 - \theta^3$

31. $f(\theta) = \sin\theta - \cos\theta$

32. $f(x) = 1,100$

In Exercises 33-38, find the equations of the tangent and normal lines to the graph of the function at the given point.

33. $f(x) = x^3 - x \text{ at } x = 1$

34. $f(t) = e^t \text{ at } t = 0$

35. $g(x) = \ln x \text{ at } t = 0$

36. $f(x) = 4\sin x \text{ at } x = \pi/2$

37. $f(x) = -2\cos x \text{ at } x = \pi/4$

38. $f(x) = 2x + 3 \text{ at } x = 5$

Review

39. Given that $e^0 = 1$, approximate the value of $e^{0.1}$ using the tangent line to $f(x) = e^x$ at $x = 0$.

40. Approximate the value of $(3.01)^4$ using the tangent line to $f(x) = x^4$ at $x = 3$.

2.4: The Product and Quotient Rules

Terms and Concepts

1. T/F: The Product Rule states that $fracddx\left(x^2 \sin x\right) = 2x \cos x$.

2. T/F: The Quotient Rule states that $\frac{d}{dx}\left(\frac{x^2}{\sin x}\right) = \frac{\cos x}{2x}$.

3. T/F: The derivatives of the trigonometric functions that start with "c" have minus signs in them.

4. What derivative rule is used to extend the Power Rule to include negative integer exponents?

5. T/F: Regardless of the function, there is always exactly one right way of computing its derivative.

6. In your own words, explain what it means to make your answers "clear."

Problems

In Exercises 7-10:
(a) Use the Product Rule to differentiate the function.
(b) Manipulate the function algebraically and differentiate without the Product Rule.
(c) Show that the answers from (a) and (b) are equivalent.

7. $f(x) = x(x^2 + 3x)$

8. $g(x) = 2x^2(5x^3)$

9. $h(s) = (2s - 1)(s + 4)$

10. $f(x) = (x^2 + 5)(3 - x^3)$

In Exercises 11-14:
(a) Use the Quotient Rule to differentiate the function.
(b) Manipulate the function algebraically and differentiate without the Quotient Rule.
(c) Shows that the answers from (a) and (b) are equivalent.

11. $f(x) = \frac{x^2+3}{x}$

12. $g(x) = \frac{x^3-2x^2}{2x^2}$

13. $h(s) = \frac{3}{4s^3}$

14. $f(t) = \frac{t^2-1}{t+1}$

In Exercises 15-29, compute the derivative of the given function.

15. $f(x) = x \sin x$

16. $f(t) = \frac{1}{t^2}(\csc t - 4)$

17. $g(x) = \frac{x+7}{x-5}$

18. $g(t) = \frac{t^5}{\cos t - 2t^2}$

19. $h(x) = \cot x - e^x$

20. $h(t) = 7t^2 + 6t - 2$

21. $f(x) = \frac{x^4+2x^3}{x^2}$

22. $f(x) = (16x^3 + 24x^2 + 3x)\frac{7x-1}{16x^3+24x^2+3x}$

23. $f(t) = t^5(\sec t + e^t)$

24. $f(x) = \frac{\sin x}{\cos x + 3}$

25. $g(x)=e^2\left(\sin(\pi/4)-1\right)$

26. $g(t)=4t^3e^t-\sin t\cos t$

27. $h(t)=\frac{t^2\sin t+3}{t^2\cos t+2}$

28. $f(x)=x^2e^x\tan x$

29. $g(x)=2x\sin x\sec x$

In Exercises 30-33, find the equations of the tangent and normal lines to the graph of g at the indicated point.

30. $g(s)=e^s(s^2+2)$ at $(0,2)$.

31. $g(t)=t\sin t$ at $\left(\frac{3\pi}{2},\frac{-3\pi}{2}\right)$

32. $g(x)=\frac{x^2}{x-1}$ at $(2,4)$.

33. $g(\theta)=\frac{\cos\theta-8\theta}{\theta+1}$ at $(0,-5)$

In Exercises 34-37, find the x-values where the graph of the function has a horizontal tangent line.

34. $f(x)=6x^2-18x-24$

35. $f(x)=x\sin x$ on $[-1,1]$

36. $f(x)=\frac{x}{x+1}$

37. $f(x)=\frac{x^2}{x+1}$

In Exercises 38-41, find the requested derivative.

38. $f(x)=x\sin x;$ find $f''(x)$.

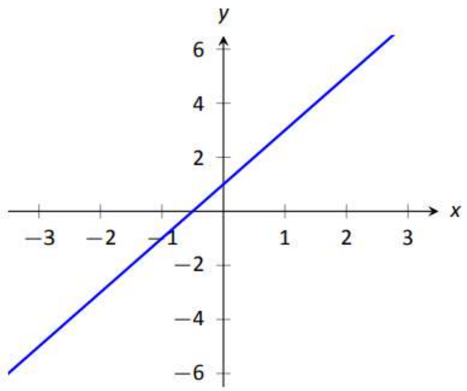
39. $f(x)=x\sin x;$ find $f^{(4)}(x)$.

40. $f(x)=\csc x;$ find $f''(x)$.

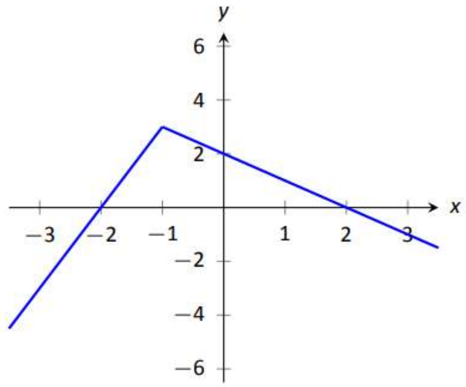
41. $f(x)=(x^3-5x+2)(x^2+x-7);$ find $f^{(8)}(x)$.

In Exercises 42-45, use the graph of $f(x)$ to sketch $f'(x)$.

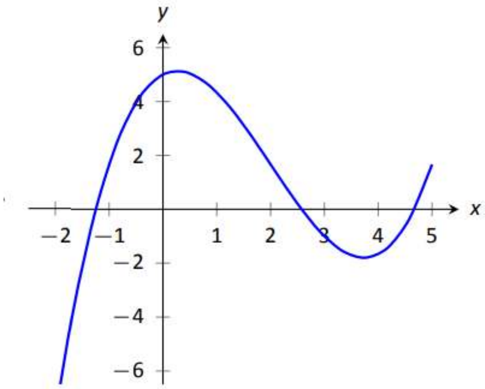
42.



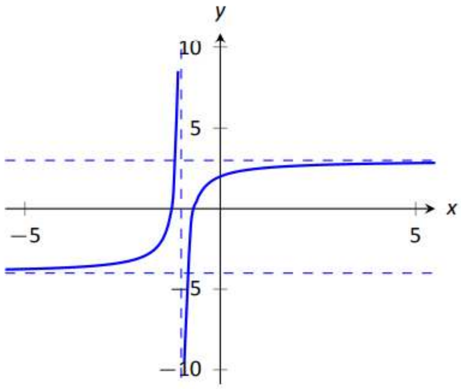
43.



44.



45.



2.5: The Chain Rule

Terms and Concepts

1. T/F: The Chain Rule describes how to evaluate the derivative of a composition of functions.

2. T/F: The Generalized Power Rule states that $\frac{d}{dx}\left(g(x)^n\right) = n\left(g(x)\right)^{n-1}$.

3. T/F: $\frac{d}{dx}\left(\ln(x^2)\right) = \frac{1}{x^2}$.

4. T/F: $\frac{d}{dx}(3^x) \approx 1.1 \cdot 3^x$.

5. T/F: $\frac{dx}{dy} = \frac{dx}{dt} \cdot \frac{dt}{dy}$

6. T/F: Taking the derivative of $f(x)=x^2\sin(5x)$ requires the use of both the Product and Chain Rules.

Problems

In Exercises 7-28, compute the derivatives of the given function.

7. $f(x)=\left(4x^3-x\right)^{10}$

8. $f(t)=(3t-2)^5$

9. $g(\theta)=(\sin\theta+\cos\theta)^3$

10. $h(t)=e^{3t^2+y-1}$

11. $f(x)=\left(x+\frac{1}{x}\right)^4$

12. $f(x)=\cos(3x)$

13. $g(x)=\tan(5x)$

14. $h(t)=\sin^4(2t)$

15. $p(t)=\cos^3(t^2+3t+1)$

16. $f(x)=\ln(\cos x)$

17. $f(x)=\ln(x^2)$

18. $f(x)=2\ln(x)$

19. $g(r)=4^r$

20. $g(t)=5^{\cos t}$

21. $g(t)=15^2$

22. $m(w)=\dfrac{3^w}{2^w}$

23. $h(t)=\dfrac{2^t+3}{3^t+2}$

24. $m(w)=\dfrac{3^w+1}{2^w}$

25. $f(x)=\dfrac{3^{x^2}+x}{2^{x^2}}$

26. $f(x)=x^2\sin(5x)$

27. $g(t)=\cos(t^2+3t)\sin(5t-7)$

28. $g(t)=\cos(\frac{1}{t})e^{5t^2}$

In Exercises 29-32, find the equation of tangent and normal lines to the graph of the function at the given point. Note: the functions here are the same as in Exercises 7 through 10.

29. $f(x)=\left(4x^3-x\right)^{10}$ at $x=0$

30. $f(t)=(3t-2)^5$ at $t=1$

31. $g(\theta)=(\sin\theta+\cos\theta)^3$ at $\theta=\pi/2$

32. $h(t)=e^{3t^2+t-1}$ at $t=-1$

33. Compute $\frac{d}{dx}\left(\ln(kx)\right)$ two ways:
(a) Using the Chain rule, and
(b) by first using the logarithm rule $\ln(ab)=\ln a+\ln b$, then taking the derivative.

34. Compute $\frac{d}{dx}\left(\ln(x^k)\right)$ two ways:
(a) Using the Chain Rule, and
(b) by first using the logarithm rule $\ln(a^p)=p\ln a$, then taking the derivative.

Review

35. The “wind chill factor” is a measurement of how cold it “feels” during cold, windy weather. Let $W(w)$ be the wind chill factor, in degrees Fahrenheit, when it is 25 F outside with a wind of w mph.
(a) What are the units of $W'(w)$?
(b) What would you expect the sign of $W'(10)$ to be?

36. Find the derivatives of the following functions.
(a) $f(x) = x^2 e^x \cot x$
(b) $g(x) = 2^x 3^x 4^x$

2.6: Implicit Differentiation

Terms and Concepts

1. In your own words, explain the difference between implicit functions and explicit functions.

2. Implicit differentiation is based on what other differentiation rule?

3. T/F: Implicit differentiation can be used to find the derivative of $y = \sqrt{x}$.

4. T/F: Implicit differentiation can be used to find the derivative of $y = x^{3/4}$.

Problems

In Exercises 5-12, compute the derivative of the given function.

5. $f(x) = \sqrt{x} + \frac{1}{\sqrt{x}}$

6. $f(x) = \sqrt[3]{x} + x^{2/3}$

7. $f(x) = \sqrt{1 - t^2}$

8. $g(t) = \sqrt{t} \sin t$

9. $h(x) = x^{1.5}$

10. $f(x) = x^{\pi} + x^{1.9} + \pi^{1.9}$

11. $g(x) = \frac{x+7}{\sqrt{x}}$

12. $f(t) = \sqrt[5]{t}\,(\sec t + e^t)$

In Exercises 13-25, find $\frac{dy}{dx}$ using implicit differentiation.

13. $x^4 + y^2 + y = 7$

14. $x^{2/5} + y^{2/5} = 1$

15. $\cos(x) + \sin(y) = 1$

16. $fracxy = 10$

17. $\frac{y}{x} = 10$

18. $x^2 e^2 + 2^y = 5$

19. $x^2 \tan y = 50$

20. $(3x^2 + 2y^3)^4 = 2$

21. $(y^2 + 2y - x)^2 = 200$

22. $\frac{x^2+y}{x+y^2} = 17$

23. $\frac{\sin(x)+y}{\cos(y)+x} = 1$

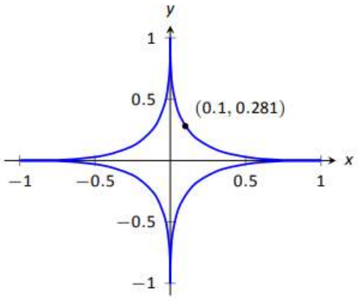
24. $\ln(x^2 + y^2) = e$

25. $\ln(x^2 + xy + y^2) = 1$

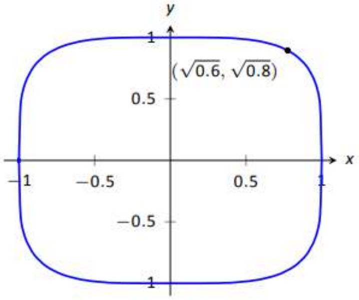
26. Show that $\frac{dy}{dx}$ is the same for each of the following implicitly defined functions.
(a) $xy = 1$
(b) $x^2y^2 = 1$
(c) $\sin(xy) = 1$
(d) $\ln(xy) = 1$

In Exercises 27-31, find the equation of the tangent line to the graph of the implicitly defined function at the indicated points. As a visual aid, each function is graphed.

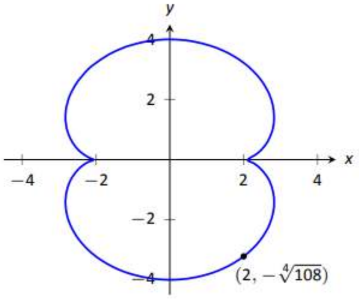
27. $x^{2/5} + y^{2/5} = 1$
(a) At (1,0)
(b) At (0.1, 0.281) (which does not *exactly* lie on the curve, but is very close).



28. $x^4 + y^4 = 1$
(a) At (1,0).
(b) At $(\sqrt{0.6}, \sqrt{0.8})$.
(c) At (0,1).

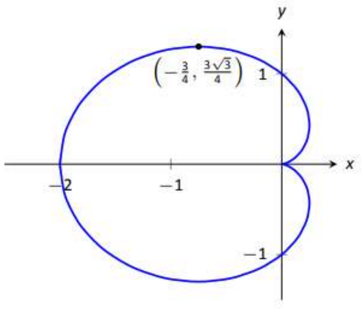


29. $(x^2 + y^2 - 4)^3 = 108y^2$
(a) At (0,4).
(b) At $(2, -\sqrt[4]{108})$

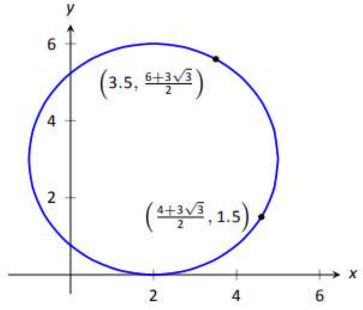


30. $(x^2 + y^2 + x)^2 = x^2 + y^2$
(a) At (0,1).

(b) At $\left(-\frac{3}{4}, \frac{3\sqrt{3}}{4}\right)$.



31. $(x-2)^2 + (y-3)^2 = 9$
(a) At $\left(\frac{7}{2}, \frac{6+3\sqrt{3}}{2}\right)$.
(b) At $\left(\frac{4+3\sqrt{3}}{2}, \frac{3}{2}\right)$.



In Exercises 32-35, an implicitly defined function is given. Find $\frac{d^2y}{dx^2}$. Note: these are the same problems used in Exercises 13-16.

32. $x^4 + y^2 + y = 7$

33. $x^{2/5} + y^{2/5} = 1$

34. $\cos x + \sin y = 1$

35. $\frac{x}{y} = 10$

In Exercises 36-41, use logarithmic differentiation to find $\frac{dy}{dx}$, then find the equation of the tangent line at the indicated x-value.

36. $y = (1+x)^{1/x}, \quad x = 1$

37. $y = 2x^{x^2}, \quad x = 1$

38. $y = \frac{x^x}{x+1}, \quad x = 1$

39. $y = x^{\sin(x)+2}, \quad x = 1$

40. $y = \frac{x+1}{x+2}, \quad x = 1$

41. $y = \frac{(x+1)(x+2)}{(x+3)(x+4)}, \quad x = 1$

2.7: Derivatives of Inverse Functions

Terms and Concepts

1. T/F: Every function has an inverse.

2. In your own words explain what it means for a function to be "one on one."

3. If (1,10) lies on the graph of $y = f(x)$, what can be said about the graph of $y = f^{-1}(x)$?

4. If (1,10) lies on the graph of $y=f(x)$ and $f'(1)=5$, what can be said about $y=f^{-1}(x)$?

Problems

In Exercises 5-8, verify that the given functions are inverses.

5. $f(x)=2x+6$ and $g(x)=\frac{1}{2}x-3$

6. $f(x)=x^2+6x+11,\ x\geq 3$ and $g(x)=\sqrt{x-2}-3,\ x\geq 2$

7. $f(x)=\frac{3}{x-5},\ x\neq 5$ and $g(x)=\frac{3+5x}{x},\ x\neq 0$

8. $f(x)=\frac{x+1}{x-1},\ x\neq 1$ and $g(x)=f(x)$

In Exercises 9-14, an invertible function $f(x)$ is given along with a point that lies on its graph. Using Theorem 22, evaluate $\left(f^{-1}\right)'(x)$ at the indicated value.

9. $f(x)=5x+10$
Point = (2,20)
Evaluate $\left(f^{-1}\right)'(20)$

10. $f(x)=x^2-2x+4,\ x\geq 1$
Point = $(3,7)$
Evaluate $\left(f^{-1}\right)'(7)$

11. $f(x)=\sin 2x,\ -\pi/4\leq x\leq \pi/4$
Point = $(\pi/6,\sqrt{3}/2)$
Evaluate $\left(f^{-1}\right)'(\sqrt{3}/2)$

12. $f(x)=x^3-6x^2+15x-2$
Point = $(1,8)$
Evaluate $\left(f^{-1}\right)'(8)$

13. $f(x)=\frac{1}{1+x^2},\ x\geq 0$
Point = $(1,1/2)$
Evaluate $\left(f^{-1}\right)'(1/2)$

14. $f(x)=6e^{3x}$
Point = $(0,6)$
Evaluate $\left(f^{-1}\right)'(6)$

In Exercises 15-24, compute the derivative of the given function.

15. $h(t)=\sin^{-1}(2t)$

16. $f(t)=\sec^{-1}(2t)$

17. $g(x)=\tan^{-1}(2x)$

18. $f(x)=x\sin^{-1}(x)$

19. $g(t)=\sin t\cos^{-1}t$

20. $f(t)=\ln te^t$

21. $h(x)=\frac{\sin^{-1}x}{\cos^{-1}x}$

22. $g(x)=\tan^{-1}(\sqrt{x})$

23. $f(x)=\sec^{-1}(1/x)$

24. $f(x)=\sin(\sin^{-1}x)$

In Exercises 25-27, compute the derivative of the given function in two ways:
(a) By simplifying first, then taking the derivative, and
(b) by using the Chain Rule first then simplifying.

Very that the two answers are the same.

25. $f(x)=\sin\left(\sin^{-1}x\right)$

26. $f(x)=\tan^{-1}(\tan x)$

27. $f(x)=\sin\left(\cos^{-1}x\right)$

In Exercises 28-29, find the equation of the line tangent to the graph of f at the indicated value.

28. $f(x)=\sin^{-1}x \text{ at } x=\frac{\sqrt{2}}{2}$

29. $f(x)=\cos^{-1}(2x) \text{ at } x=\frac{\sqrt{3}}{4}$

Review

30. Find $\frac{dy}{dx}$, where $x^2y-y^2x=1$.

31. Find the equation of the line tangent to the graph of $x^2+y^2+xy=7$ at the point (1,2).

32. Let $f(x)=x^3+x$. Evaluate $\lim\limits_{s\to 0}\frac{f(x+s)-f(x)}{s}$.

2.E: Applications of Derivatives(Exercises) is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

3: The Graphical Behavior of Functions

Our study of limits led to continuous functions, which is a certain class of functions that behave in a particularly nice way. Limits then gave us an even nicer class of functions, functions that are differentiable. This chapter explores many of the ways we can take advantage of the information that continuous and differentiable functions provide.

3.1: Extreme Values
3.2: The Mean Value Theorem
3.3: Increasing and Decreasing Functions
3.4: Concavity and the Second Derivative
3.5: Curve Sketching
3.E: Applications of the Graphical Behavior of Functions(Exercises)

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 3: The Graphical Behavior of Functions is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

3.1: Extreme Values

Our study of limits led to continuous functions, which is a certain class of functions that behave in a particularly nice way. Limits then gave us an even nicer class of functions, functions that are differentiable.

This chapter explores many of the ways we can take advantage of the information that continuous and differentiable functions provide.

Given any quantity described by a function, we are often interested in the largest and/or smallest values that quantity attains. For instance, if a function describes the speed of an object, it seems reasonable to want to know the fastest/slowest the object traveled. If a function describes the value of a stock, we might want to know how the highest/lowest values the stock attained over the past year. We call such values *extreme values*.

Definition $3.1.1$: Minima and Maxima

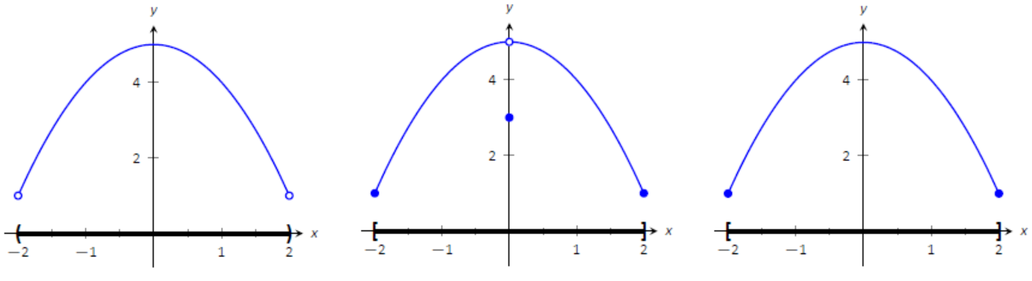
Let f be defined on an interval I containing c.

- $f(c)$ is the **minimum** (also, **absolute minimum**) of f on I if $f(c) \leq f(x)$ for all x in I.
- $f(c)$ is the **maximum**} (also, **absolute maximum**) of f on I if $f(c) \geq f(x)$ for all x in I.

The maximum and minimum values are the **extreme values**, or **extrema**, of f on I.

The extreme values of a function are "y" values, values the function attains, not the input values.

Consider Figure $3.1.1$. The function displayed in (a) has a maximum, but no minimum, as the interval over which the function is defined is open. In (b), the function has a minimum, but no maximum; there is a discontinuity in the "natural" place for the maximum to occur. Finally, the function shown in (c) has both a maximum and a minimum; note that the function is continuous and the interval on which it is defined is closed.



(a) (b) (c)

Figure $3.1.1$: Graphs of functions with and without extreme values

It is possible for discontinuous functions defined on an open interval to have both a maximum and minimum value, but we have just seen examples where they did not. On the other hand, continuous functions on a closed interval *always* have a maximum and minimum value.

Theorem $3.1.1$: The Extreme Value Theorem

Let f be a continuous function defined on a closed interval I. Then f has both a maximum and minimum value on I.

This theorem states that f has extreme values, but it does not offer any advice about how/where to find these values. The process can seem to be fairly easy, as the next example illustrates. After the example, we will draw on lessons learned to form a more general and powerful method for finding extreme values.

Example $3.1.1$: Approximating Extreme Values

Consider $f(x) = 2x^3 - 9x^2$ on $I = [-1, 5]$, as graphed in Figure $3.1.2$. Approximate the extreme values of f.

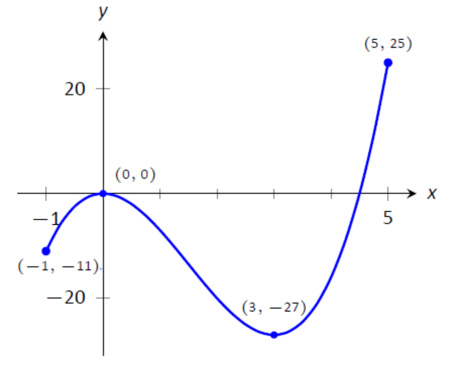


Figure $3.1.2$: A graph of $f(x)=2x^3-9x^2$ as in Example $3.1.1$

Solution

The graph is drawn in such a way to draw attention to certain points. It certainly seems that the smallest y value is -27, found when $x=3$. It also seems that the largest y value is 25, found at the endpoint of I, $x=5$. We use the word *seems*, for by the graph alone we cannot be sure the smallest value is not less than -27. Since the problem asks for an approximation, we approximate the extreme values to be 25 and -27.

Notice how the minimum value came at "the bottom of a hill," and the maximum value came at an endpoint. Also note that while 0 is not an extreme value, it would be if we narrowed our interval to $[-1,4]$. The idea that the point $(0,0)$ is the location of an extreme value for some interval is important, leading us to a definition.

Local and Relative Extrema

The terms *local minimum* and *local maximum* are often used as synonyms for *relative minimum* and *relative maximum.* We briefly practice using these definitions.

Definition $3.1.2$: Relative Minimum and Relative Maximum

Let f be defined on an interval I containing c.

1. If there is an open interval containing c such that $f(c)$ is the minimum value, then $f(c)$ is a **relative minimum** of f. We also say that f has a relative minimum at $(c,f(c))$.
2. If there is an open interval containing c such that $f(c)$ is the maximum value, then $f(c)$ is a **relative maximum** of f. We also say that f has a relative maximum at $(c,f(c))$.

The relative maximum and minimum values comprise the **relative extrema** of f.

Example $3.1.2$: Approximating Relative Extrema

Consider $f(x)=(3x^4-4x^3-12x^2+5)/5$, as shown in Figure $3.1.3$. Approximate the relative extrema of f. At each of these points, evaluate f'.

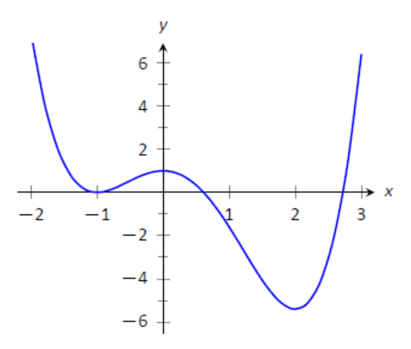


Figure $3.1.3$: A graph of $f(x) = (3x^4 - 4x^3 - 12x^2 + 5)/5$ as in Example $3.1.2$.

Solution

We still do not have the tools to exactly find the relative extrema, but the graph does allow us to make reasonable approximations. It seems f has relative minima at $x = -1$ and $x = 2$, with values of $f(-1) = 0$ and $f(2) = -5.4$. It also seems that f has a relative maximum at the point $(0, 1)$.

We approximate the relative minima to be 0 and -5.4; we approximate the relative maximum to be 1.

It is straightforward to evaluate $f'(x) = \frac{1}{5}(12x^3 - 12x^2 - 24x)$ at $x = 0, 1$ and 2. In each case, $f'(x) = 0$.

Example $3.1.3$: Approximating Relative Extrema

Approximate the relative extrema of $f(x) = (x-1)^{2/3} + 2$, shown in Figure $3.1.4$. At each of these points, evaluate f'.

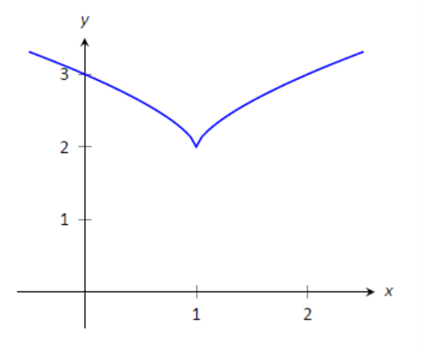


Figure $3.1.4$: A graph of $f(x) = (x-1)^{2/3} + 2$ as in Example $3.1.3$.

Solution

The figure implies that f does not have any relative maxima, but has a relative minimum at $(1, 2)$. In fact, the graph suggests that not only is this point a relative minimum, $y = f(1) = 2$ the minimum value of the function.

We compute $f'(x) = \frac{2}{3}(x-1)^{-1/3}$. When $x = 1$, f' is undefined.

What can we learn from the previous two examples? We were able to visually approximate relative extrema, and at each such point, the derivative was either 0 or it was not defined. This observation holds for all functions, leading to a definition and a theorem.

Definition $3.1.3$: Critical Numbers and Critical Points

Let f be defined at c. The value c is a **critical number** (or **critical value**) of f if $f'(c) = 0$ or $f'(c)$ is not defined.

If c is a critical number of f, then the point $(c, f(c))$ is a **critical point** of f.

Theorem $3.1.1$: Relative Extrema and Critical Points

Let a function f have a relative extrema at the point $(c, f(c))$. Then c is a critical number of f.

Be careful to understand that this theorem states "All relative extrema occur at critical points." It does not say "All critical numbers produce relative extrema." For instance, consider $f(x) = x^3$. Since $f'(x) = 3x^2$, it is straightforward to determine that $x = 0$ is a critical number of f. However, f has no relative extrema, as illustrated in Figure $3.1.5$.

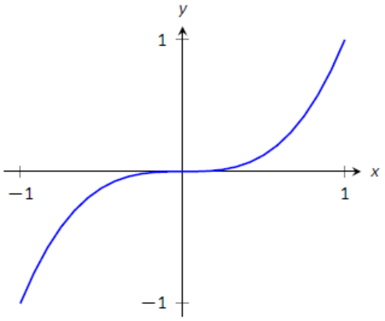


Figure 3.1.5: A graph of $f(x)=x^3$ which has a critical value of $x=0$, but no relative extrema.

Theorem 3.1.1 states that a continuous function on a closed interval will have absolute extrema, that is, both an absolute maximum and an absolute minimum. These extrema occur either at the endpoints or at critical values in the interval. We combine these concepts to offer a strategy for finding extrema.

Key Idea 2: Finding Extrema on a Closed Interval

Let f be a continuous function defined on a closed interval $[a,b]$. To find the maximum and minimum values of f on $[a,b]$

1. Evaluate f at the endpoints a and b of the interval.
2. Find the critical numbers of f in $[a,b]$.
3. Evaluate f at each critical number.
4. The absolute maximum of f is the largest of these values, and the absolute minimum of f is the least of these values.

We practice these ideas in the next examples.

Example 3.1.4: Finding Extreme Values

Find the extreme values of $f(x)=2x^3+3x^2-12x$ on $[0,3]$, graphed in Figure 3.1.6.

We follow the steps outlined in the Key Idea 2. We first evaluate f at the endpoints:

$$f(0)=0 \quad \text{and} \quad f(3)=45. \tag{3.1.1}$$

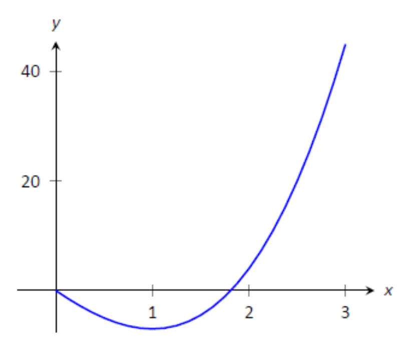


Figure 3.1.6: A graph of $f(x)=2x^3+3x^2-12x$ on $[0,3]$ as in Example 3.1.4.

Next, we find the critical values of f on $[0,3]$. $f'(x)=6x^2+6x-12=6(x+2)(x-1)$; therefore the critical values of f are $x=-2$ and $x=1$. Since $x=-2$ does not lie in the interval $[0,3]$, we ignore it. Evaluating f at the only critical number in our interval gives: $f(1)=-7$.

Table 3.1.1 gives f evaluated at the "important" x values in $[0,3]$. We can easily see the maximum and minimum values of f: the maximum value is 45 and the minimum value is -7.

Table 3.1.1: Finding the extreme values of f in Example 3.1.4

x	$f(x)$

x	$f(x)$
0	0
1	-7
3	45

Note that all this was done without the aid of a graph; this work followed an analytic algorithm and did not depend on any visualization. Figure $3.1.6$ shows f and we can confirm our answer, but it is important to understand that these answers can be found without graphical assistance.

We practice again.

Example $3.1.5$: Finding Extreme Values

Find the maximum and minimum values of f on $[-4,2]$, where

$$f(x)=\left\{\begin{array}{cc}(x-1)^2 & x\leq 0\\ x+1 & x>0\end{array}\right. . \tag{3.1.2}$$

Solution

Here f is piecewise--defined, but we can still apply the Key Idea 2. Evaluating f at the endpoints gives:

$$f(-4)=25 \quad \text{and} \quad f(2)=3. \tag{3.1.3}$$

We now find the critical numbers of f. We have to define f' in a piecewise manner; it is

$$f'(x)=\left\{\begin{array}{cc}2(x-1) & x<0\\ 1 & x>0\end{array}\right. . \tag{3.1.4}$$

Note that while f is defined for all of $[-4,2]$, f' is not, as the derivative of f does not exist when $x=0$. (From the left, the derivative approaches -2; from the right the derivative is 1.) Thus one critical number of f is $x=0$.

We now set $f'(x)=0$. When $x>0$, $f'(x)$ is never 0. When $x<0$, $f'(x)$ is also never 0. (We may be tempted to say that $f'(x)=0$ when $x=1$. However, this is nonsensical, for we only consider $f'(x)=2(x-1)$ when $x<0$, so we will ignore a solution that says $x=1$.)

So we have three important x values to consider: $x=-4,2$ and 0. Evaluating f at each gives, respectively, 25, 3 and 1, shown in Table $3.1.2$. Thus the absolute minimum of f is 1; the absolute maximum of f is 25. Our answer is confirmed by the graph of f in Figure $3.1.7$.

Table $3.1.2$: Finding the extreme values of f in Example $3.1.5$

x	$f(x)$
-4	25
0	1
2	3

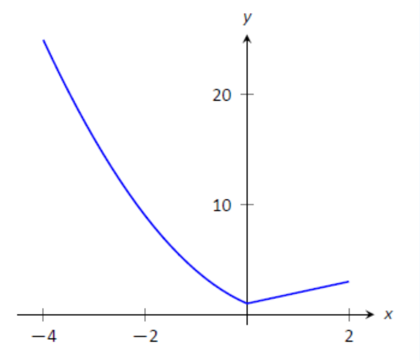


Figure $3.1.7$: A graph of $f(x)$ on $[-4,2]$ as in Example $3.1.5$.

Example $3.1.6$: Finding Extreme Values

Find the extrema of $f(x)=\cos(x^2)$ on $[-2,2]$.

Solution

We again use Key Idea 3. Evaluating f at the endpoints of the interval gives: $f(-2)=f(2)=\cos(4)\approx -0.6536.$ We now find the critical values of f.

Applying the Chain Rule, we find $f'(x)=-2x\sin(x^2)$. Set $f'(x)=0$ and solve for x to find the critical values of f.

We have $f'(x)=0$ when $x=0$ and when $\sin(x^2)=0$. In general, $\sin t=0$ when $t=\ldots-2\pi,-\pi,0,\pi,\ldots$ Thus $\sin(x^2)=0$ when $x^2=0,\pi,2\pi,\ldots$ (x^2 is always positive so we ignore $-\pi$, etc.) So $\sin(x^2)=0$ when $x=0,\pm\sqrt{\pi},\pm\sqrt{2\pi},\ldots$. The only values to fall in the given interval of $[-2,2]$ are $-\sqrt{\pi}$ and $\sqrt{\pi}$, approximately ± 1.77.

We again construct a table of important values in Table $3.1.3$. In this example we have 5 values to consider: $x=0,\pm 2,\pm\sqrt{\pi}$.

Table $3.1.3$: Finding the extrema of $f(x)=\cos(x^2)$ in Example $3.1.5$

x	$f(x)$
-2	-0.65
$-\sqrt{\pi}$	-1
0	1
$\sqrt{\pi}$	-1
2	-0.65

From the table it is clear that the maximum value of f on $[-2,2]$ is 1; the minimum value is -1. The graph in Figure $3.1.8$ confirms our results.

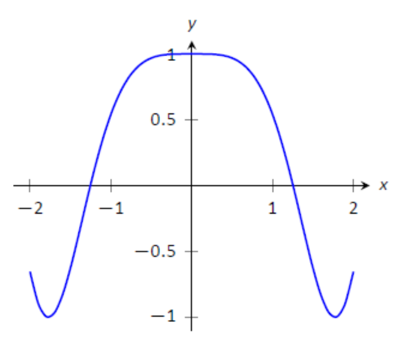


Figure $3.1.8$: A graph of $f(x)=\cos(x^2)$ on $[-2,2]$ as in Example $3.1.5$.

We consider one more example.

Example $3.1.7$: Finding Extreme Values

Find the extreme values of $f(x)=\sqrt{1-x^2}$.

Solution

A closed interval is not given, so we find the extreme values of f on its domain. f is defined whenever $1-x^2\geq 0$; thus the domain of f is $[-1,1]$. Evaluating f at either endpoint returns 0.

Using the Chain Rule, we find $f'(x)=\frac{-x}{\sqrt{1-x^2}}$. The critical points of f are found when $f'(x)=0$ or when f' is undefined. It is straightforward to find that $f'(x)=0$ when $x=0$, and f' is undefined when $x=\pm 1$, the endpoints of the interval. The table of important values is given in Table $3.1.4$.

Table $3.1.4$: Finding the extrema of the half--circle in Example $3.1.7$

x	$f(x)$
-1	0
0	1
1	0

The maximum value is 1, and the minimum value is 0.

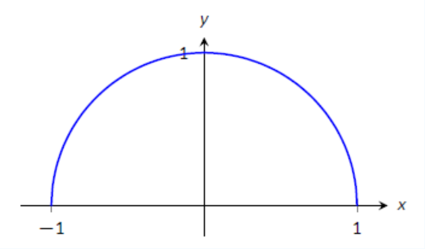


Figure $3.1.9$: A graph of $f(x)=\sqrt{1-x^2}$ on $[-1,1]$ as in Example $3.1.7$

Note: We implicitly found the derivative of $x^2+y^2=1$, the unit circle, in the section on Implicit Differentiation as $\frac{dy}{dx}=-x/y$. In Example $3.1.7$, half of the unit circle is given as $y=f(x)=\sqrt{1-x^2}$. We found $f'(x)=\frac{-x}{\sqrt{1-x^2}}$. Recognize that the denominator of this fraction is y; that is, we again found $f'(x)=\frac{dy}{dx}=-x/y$.

We have seen that continuous functions on closed intervals always have a maximum and minimum value, and we have also developed a technique to find these values. In the next section, we further our study of the information we can glean from "nice" functions with the Mean Value Theorem. On a closed interval, we can find the *average rate of change* of a function (as we did at the beginning of Chapter 2). We will see that differentiable functions always have a point at which their instantaneous rate of change is same as the *average* rate of change. This is surprisingly useful, as we'll see.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/
- Integrated by Justin Marshall.

This page titled 3.1: Extreme Values is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

3.2: The Mean Value Theorem

We motivate this section with the following question: Suppose you leave your house and drive to your friend's house in a city 100 miles away, completing the trip in two hours. At any point during the trip do you necessarily have to be going 50 miles per hour?

In answering this question, it is clear that the *average* speed for the entire trip is 50 mph (i.e. 100 miles in 2 hours), but the question is whether or not your *instantaneous* speed is ever exactly 50 mph. More simply, does your speedometer ever read exactly 50 mph?. The answer, under some very reasonable assumptions, is "yes.'"

Let's now see why this situation is in a calculus text by translating it into mathematical symbols.

First assume that the function $y = f(t)$ gives the distance (in miles) traveled from your home at time t (in hours) where $0 \leq t \leq 2$. In particular, this gives $f(0) = 0$ and $f(2) = 100$. The slope of the secant line connecting the starting and ending points $(0, f(0))$ and $(2, f(2))$ is therefore

$$\frac{\Delta f}{\Delta t} = \frac{f(2) - f(0)}{2 - 0} = \frac{100 - 0}{2} = 50\,\text{mph}. \tag{3.2.1}$$

The slope at any point on the graph itself is given by the derivative $f'(t)$. So, since the answer to the question above is "yes," this means that at some time during the trip, the derivative takes on the value of 50 mph. Symbolically,

$$f'(c) = \frac{f(2) - f(0)}{2 - 0} = 50 \tag{3.2.2}$$

for some time $0 \leq c \leq 2$.

How about more generally? Given any function $y = f(x)$ and a range $a \leq x \leq b$ does the value of the derivative at some point between a and b have to match the slope of the secant line connecting the points $(a, f(a))$ and $(b, f(b))$? Or equivalently, does the equation

$$f'(c) = \frac{f(b) - f(a)}{b - a} \tag{3.2.3}$$

have to hold for some $a < c < b$?

Let's look at two functions in an example.

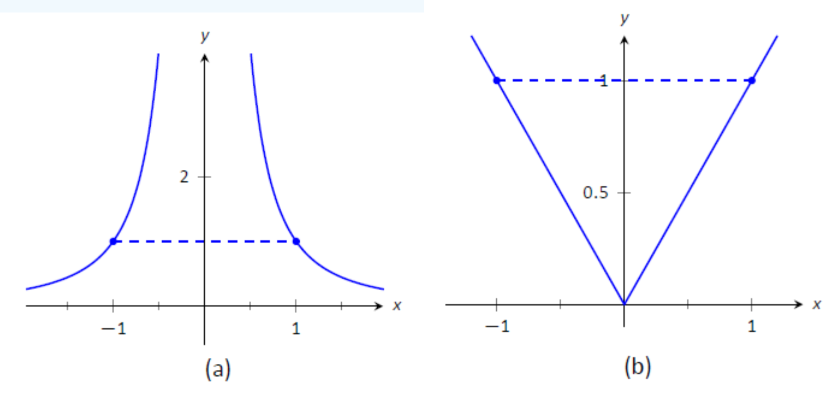
Example $3.2.1$: Comparing average and instantaneous rates of change

Consider functions

$$f_1(x)=\frac{1}{x^2}\quad \text{and} \quad f_2(x) = |x|$$

with $a = -1$ and $b = 1$ as shown in Figure $3.2.1$ (a) and (b), respectively. Both functions have a value of 1 at a and b. Therefore the slope of the secant line connecting the end points is 0 in each case. But if you look at the plots of each, you can see that there are no points on either graph where the tangent lines have slope zero. Therefore we have found that there is no c in $[-1, 1]$ such that

$$f'(c) = \frac{f(1)-f(-1)}{1-(-1)} = 0.$$



(a) (b)

Figure* $3.2.1$*: A graph of $f_1(x) = 1/x^2$ and $f_2(x) = |x|$ in Example $3.2.1$.

So what went "wrong"? It may not be surprising to find that the discontinuity of f_1 and the corner of f_2 play a role. If our functions had been continuous and differentiable, would we have been able to find that special value c? This is our motivation for the following theorem.

Theorem 3.2.1: The Mean Value Theorem of Differentiation

Let $y = f(x)$ be continuous function on the closed interval $[a, b]$ and differentiable on the open interval (a, b). There exists a value c, $a < c < b$, such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}. \tag{3.2.4}$$

That is, there is a value c in (a, b) where the instantaneous rate of change of f at c is equal to the average rate of change of f on $[a, b]$.

Note that the reasons that the functions in Example $3.2.1$ fail are indeed that f_1 has a discontinuity on the interval $[-1, 1]$ and f_2 is not differentiable at the origin.

We will give a proof of the Mean Value Theorem below. To do so, we use a fact, called Rolle's Theorem, stated here.

Theorem 3.2.2: Rolle's Theorem

Let f be continuous on $[a, b]$ and differentiable on (a, b), where $f(a) = f(b)$. There is some c in (a, b) such that $f'(c) = 0$.

Consider Figure $3.2.2$ where the graph of a function f is given, where $f(a) = f(b)$. It should make intuitive sense that if f is differentiable (and hence, continuous) that there would be a value c in (a, b) where $f'(c) = 0$; that is, there would be a relative maximum or minimum of f in (a, b). Rolle's Theorem guarantees at least one; there may be more.

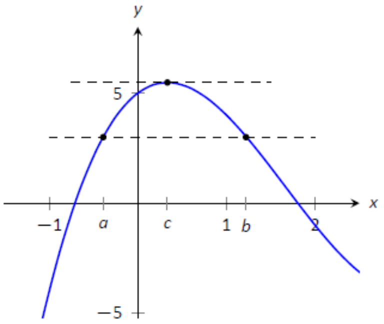


FIgure* $3.2.2$*: A graph of $f(x) = x^3 - 5x^2 + 3x + 5$, where $f(a) = f(b)$. Note the existence of c, where $a < c < b$, where $f'(c) = 0$.

Rolle's Theorem is really just a special case of the Mean Value Theorem. If $f(a) = f(b)$, then the *average* rate of change on (a, b) is 0, and the theorem guarantees some c where $f'(c) = 0$. We will prove Rolle's Theorem, then use it to prove the Mean Value Theorem.

Proof of Rolle's Theorem

Let f be differentiable on (a, b) where $f(a) = f(b)$. We consider two cases.

Case 1: Consider the case when f is constant on $[a, b]$; that is, $f(x) = f(a) = f(b)$ for all x in $[a, b]$. Then $f'(x) = 0$ for all x in $[a, b]$, showing there is at least one value c in (a, b) where $f'(c) = 0$.

Case 2: Now assume that f is not constant on $[a, b]$. The Extreme Value Theorem guarantees that f has a maximal and minimal value on $[a, b]$, found either at the endpoints or at a critical value in (a, b). Since $f(a) = f(b)$ and f is not constant, it is clear that the maximum and minimum cannot *both* be found at the endpoints. Assume, without loss of generality, that the maximum of f is not found at the endpoints. Therefore there is a c in (a, b) such that $f(c)$ is the maximum value of f. By Theorem 3.1.2, c must be a critical number of f; since f is differentiable, we have that $f'(c) = 0$, completing the proof of the theorem.

□

We can now prove the Mean Value Theorem.

Proof of the Mean Value Theorem

Define the function

$$g(x) = f(x) - \frac{f(b)-f(a)}{b-a}x.\]

We know g is differentiable on (a,b) and continuous on $[a,b]$ since f is. We can show $g(a)=g(b)$ (it is actually easier to show $g(b)-g(a)=0$, which suffices). We can then apply Rolle's theorem to guarantee the existence of $c \in (a,b)$ such that $g'(c)=0$. But note that

$$0= g'(c) = f'(c) - \frac{f(b)-f(a)}{b-a} \ ;\]

hence

$$f'(c) = \frac{f(b)-f(a)}{b-a},\]

which is what we sought to prove.

□

Going back to the very beginning of the section, we see that the only assumption we would need about our distance function $f(t)$ is that it be continuous and differentiable for t from 0 to 2 hours (both reasonable assumptions). By the Mean Value Theorem, we are guaranteed a time during the trip where our instantaneous speed is 50 mph. This fact is used in practice. Some law enforcement agencies monitor traffic speeds while in aircraft. They do not measure speed with radar, but rather by timing individual cars as they pass over lines painted on the highway whose distances apart are known. The officer is able to measure the *average* speed of a car between the painted lines; if that average speed is greater than the posted speed limit, the officer is assured that the driver exceeded the speed limit at some time.

Note that the Mean Value Theorem is an *existence* theorem. It states that a special value c *exists*, but it does not give any indication about how to find it. It turns out that when we need the Mean Value Theorem, existence is all we need

Example $3.2.2$: Using the Mean Value Theorem

Consider $f(x)=x^3+5x+5$ on $[-3,3]$. Find c in $[-3,3]$ that satisfies the Mean Value Theorem.

Solution

The average rate of change of f on $[-3,3]$ is:

$$\frac{f(3)-f(-3)}{3-(-3)} = \frac{84}{6} = 14. \tag{3.2.5}$$

We want to find c such that $f'(c)=14$. We find $f'(x)=3x^2+5$. We set this equal to 14 and solve for x.

$$\begin{aligned} f'(x) &= 14 \\ 3x^2+5 &= 14 \\ x^2 &= 3 \\ x &= \pm\sqrt{3} \approx \pm 1.732 \end{aligned}$$

We have found 2 values c in $[-3,3]$ where the instantaneous rate of change is equal to the average rate of change; the Mean Value Theorem guaranteed at least one. In Figure $3.2.3$ f is graphed with a dashed line representing the average rate of change; the lines tangent to f at $x=\pm\sqrt{3}$ are also given. Note how these lines are parallel (i.e., have the same slope) as the dashed line.

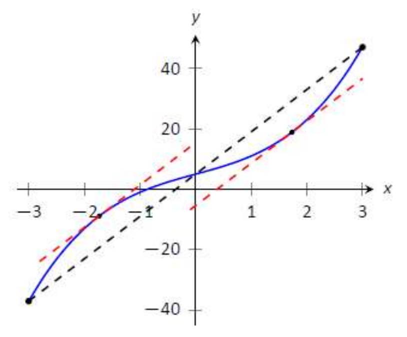


Figure $3.2.3$***:*** *Demonstrating the Mean Value Theorem in Example* $3.2.2$*.*

While the Mean Value Theorem has practical use (for instance, the speed monitoring application mentioned before), it is mostly used to advance other theory. We will use it in the next section to relate the shape of a graph to its derivative.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/
- Integrated by Justin Marshall.

This page titled 3.2: The Mean Value Theorem is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

3.3: Increasing and Decreasing Functions

Our study of "nice" functions f in this chapter has so far focused on individual points: points where f is maximal/minimal, points where $f'(x)=0$ or f' does not exist, and points c where $f'(c)$ is the average rate of change of f on some interval.

In this section we begin to study how functions behave *between* special points; we begin studying in more detail the shape of their graphs.

We start with an intuitive concept. Given the graph in Figure $3.3.1$, where would you say the function is *increasing*? *Decreasing*?

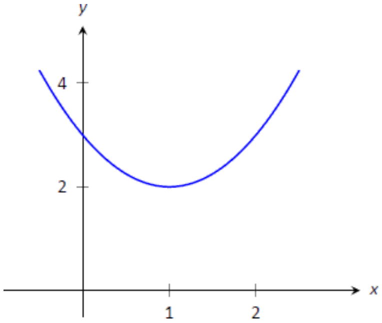


***Figure** $3.3.1$: A graph of a function f used to illustrate the concepts of increasing and decreasing.*

Even though we have not defined these terms mathematically, one likely answered that f is increasing when $x>1$ and decreasing when $x<1$. We formally define these terms here.

Definition: Increasing and Decreasing Functions

Let f be a function defined on an interval I.\index{increasing function}\index{decreasing function}\index{increasing function!strictly}\index{decreasing function!strictly}

1. f is **increasing** on I if for every $a<b$ in I, $f(a)\leq f(b)$.
2. f is **decreasing** on I if for every $a<b$ in I, $f(a)\geq f(b)$.

A function is **strictly increasing** when $a<b$ in I implies $f(a)<f(b)$, with a similar definition holding for **strictly decreasing**.

Informally, a function is increasing if as x gets larger (i.e., looking left to right) $f(x)$ gets larger.

Our interest lies in finding intervals in the domain of f on which f is either increasing or decreasing. Such information should seem useful. For instance, if f describes the speed of an object, we might want to know when the speed was increasing or decreasing (i.e., when the object was accelerating vs. decelerating). If f describes the population of a city, we should be interested in when the population is growing or declining.

To find such intervals, we again consider secant lines. Let f be an increasing, differentiable function on an open interval I, such as the one shown in Figure $3.3.2$, and let $a<b$ be given in I. The secant line on the graph of f from $x=a$ to $x=b$ is drawn; it has a slope of $(f(b)-f(a))/(b-a)$. But note:

$$\begin{gathered}\frac{f(b)-f(a)}{b-a} \Rightarrow \frac{\text{numerator } >0}{\text{denominator } >0} \Rightarrow \text{slope of the secent line} >0 \\ \Rightarrow \text{Average rate of chjange of } f \text{ on } [a,b] \text{ is } >0.\end{gathered} \tag{3.3.1}$$

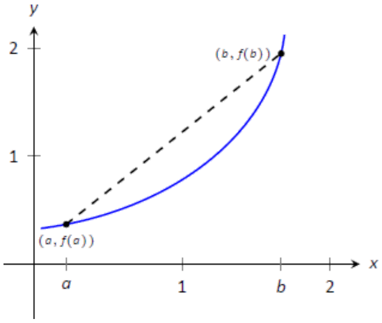


Figure* $3.3.2$*: Examining the secant line of an increasing function.

We have shown mathematically what may have already been obvious: when f is increasing, its secant lines will have a positive slope. Now recall the Mean Value Theorem guarantees that there is a number c, where $a < c < b$, such that

$$f'(c) = \frac{f(b) - f(a)}{b - a} > 0. \tag{3.3.2}$$

By considering all such secant lines in I, we strongly imply that $f'(x) \geq 0$ on I. A similar statement can be made for decreasing functions.

Our above logic can be summarized as "If f is increasing, then f' is probably positive." Theorem $3.3.1$ below turns this around by stating "If f' is postive, then f is increasing." This leads us to a method for finding when functions are increasing and decreasing.

THeorem $3.3.1$: Test For Increasing/Decreasing Functions

Let f be a continuous function on $[a, b]$ and differentiable on (a, b).

1. If $f'(c) > 0$ for all c in (a, b), then f is increasing on $[a, b]$.
2. If $f'(c) < 0$ for all c in (a, b), then f is decreasing on $[a, b]$.
3. If $f'(c) = 0$ for all c in (a, b), then f is constant on $[a, b]$.

Note: Theorem $3.3.1$ also holds if $f'(c) = 0$ for a finite number of values of c in I.

Let a and b be in I where $f'(a) > 0$ and $f'(b) < 0$. It follows from the Intermediate Value Theorem that there must be some value c between a and b where $f'(c) = 0$. This leads us to the following method for finding intervals on which a function is increasing or decreasing.

Key Idea 3: Finding Intervals on Which f is Increasing or Decreasing

Let f be a differentiable function on an interval I. To find intervals on which f is increasing and decreasing:

1. Find the critical values of f. That is, find all c in I where $f'(c) = 0$ or f' is not defined.
2. Use the critical values to divide I into subintervals.
3. Pick any point p in each subinterval, and find the sign of $f'(p)$.
    a. If $f'(p) > 0$, then f is increasing on that subinterval.
    b. If $f'(p) < 0$, then f is decreasing on that subinterval.

We demonstrate using this process in the following example.

Example $3.3.1$: Finding intervals of increasing/decreasing

Let $f(x) = x^3 + x^2 - x + 1$. Find intervals on which f is increasing or decreasing.

Solution

Using the Key Idea 3, we first find the critical values of f. We have $f'(x) = 3x^2 + 2x - 1 = (3x - 1)(x + 1)$, so $f'(x) = 0$ when $x = -1$ and when $x = 1/3$. f' is never undefined.

Since an interval was not specified for us to consider, we consider the entire domain of f which is $(-\infty, \infty)$. We thus break the whole real line into three subintervals based on the two critical values we just found: $(-\infty, -1)$, $(-1, 1/3)$ and $(1/3, \infty)$. This is shown in Figure $3.3.3$.

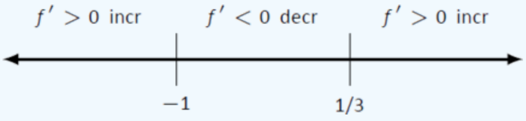


***Figure* $3.3.3$*: Number line for f in Example* $3.3.1$**

We now pick a value p in each subinterval and find the sign of $f'(p)$. All we care about is the sign, so we do not actually have to fully compute $f'(p)$; pick "nice" values that make this simple.

Subinterval 1, $(-\infty,-1)$: We (arbitrarily) pick $p=-2$. We can compute $f'(-2)$ directly: $f'(-2)=3(-2)^2+2(-2)-1=7>0$. We conclude that f is increasing on $(-\infty,-1)$.

Note we can arrive at the same conclusion without computation. For instance, we could choose $p=-100$. The first term in $f'(-100)$, i.e., $3(-100)^2$ is clearly positive and very large. The other terms are small in comparison, so we know $f'(-100)>0$. All we need is the sign.

Subinterval 2, $(-1,1/3)$: We pick $p=0$ since that value seems easy to deal with. $f'(0)=-1<0$. We conclude f is decreasing on $(-1,1/3)$.

Subinterval 3, $(1/3,\infty)$: Pick an arbitrarily large value for $p>1/3$ and note that $f'(p)=3p^2+2p-1>0$. We conclude that f is increasing on $(1/3,\infty)$.

We can verify our calculations by considering Figure $3.3.4$, where f is graphed. The graph also presents f'; note how $f'>0$ when f is increasing and $f'<0$ when f is decreasing.

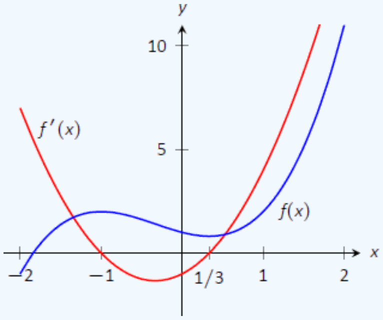


***Figure** $3.3.4$**:** A graph of $f(x)$ in Example $3.3.1$, showing where f is increasing and decreasing.*

One is justified in wondering why so much work is done when the graph seems to make the intervals very clear. We give three reasons why the above work is worthwhile.

First, the points at which f switches from increasing to decreasing are not precisely known given a graph. The graph shows us something significant happens near $x=-1$ and $x=0.3$, but we cannot determine exactly where from the graph.

One could argue that just finding critical values is important; once we know the significant points are $x=-1$ and $x=1/3$, the graph shows the increasing/decreasing traits just fine. That is true. However, the technique prescribed here helps reinforce the relationship between increasing/decreasing and the sign of f'. Once mastery of this concept (and several others) is obtained, one finds that either (a) just the critical points are computed and the graph shows all else that is desired, or (b) a graph is never produced, because determining increasing/decreasing using f' is straightforward and the graph is unnecessary.

So our second reason why the above work is worthwhile is this: once mastery of a subject is gained, one has *options* for finding needed information. We are working to develop mastery.

Finally, our third reason: many problems we face "in the real world" are very complex. Solutions are tractable only through the use of computers to do many calculations for us. Computers do not solve problems "on their own," however; they need to be taught (i.e., *programmed*) to do the right things. It would be beneficial to give a function to a computer and have it return maximum and minimum values, intervals on which the function is increasing and decreasing, the locations of relative maxima, etc. The work that we are doing here is easily programmable. It is hard to teach a computer to "look at the graph and see if it is going up or down." It is easy to teach a computer to "determine if a number is greater than or less than 0."

In Section 3.1 we learned the definition of relative maxima and minima and found that they occur at critical points. We are now learning that functions can switch from increasing to decreasing (and vice--versa) at critical points. This new understanding of increasing and decreasing creates a great method of determining whether a critical point corresponds to a maximum, minimum, or neither. Imagine a function increasing until a critical point at $x=c$, after which it decreases. A quick sketch helps confirm that $f(c)$ must be a relative maximum. A similar statement can be made for relative minimums. We formalize this concept in a theorem.

THeorem 3.3.2: First Derivative Test

Let f be differentiable on I and let c be a critical number in I.

1. If the sign of f' switches from positive to negative at c, then $f(c)$ is a relative maximum of f.
2. If the sign of f' switches from negative to positive at c, then $f(c)$ is a relative minimum of f.
3. If the sign of f' does not change at c, then $f(c)$ is not a relative extrema of f.

Example 3.3.2: Using the First Derivative Test

Find the intervals on which f is increasing and decreasing, and use the First Derivative Test to determine the relative extrema of f, where

$$f(x)=\frac{x^2+3}{x-1}. \tag{3.3.3}$$

Solution

We start by noting the domain of f: $(-\infty,1)\cup(1,\infty)$. Key Idea 3 describes how to find intervals where f is increasing and decreasing *when the domain of f is an interval*. Since the domain of f in this example is the union of two intervals, we apply the techniques of Key Idea 3 to both intervals of the domain of f.

Since f is not defined at $x=1$, the increasing/decreasing nature of f could switch at this value. We do not formally consider $x = 1$ to be a critical value of f, but we will include it in our list of critical values that we find next.

Using the Quotient Rule, we find

$$f'(x)=\frac{x^2-2x-3}{(x-1)^2}. \tag{3.3.4}$$

We need to find the critical values of f; we want to know when $f'(x)=0$ and when f' is not defined. That latter is straightforward: when the denominator of $f'(x)$ is 0, f' is undefined. That occurs when $x=1$, which we've already recognized as an important value.

Note: Strictly speaking, $x=1$ is not a critical value of f as f is not defined at $x=1$. We therefore actually apply Key Idea 3 to the intervals $(-\infty,1)$ and $(1,\infty)$. We make note of $x=1$ on the number line as we recognize that the behavior of f can change there, as it is not defined there.

$f'(x)=0$ when the numerator of $f'(x)$ is 0. That occurs when $x^2-2x-3=(x-3)(x+1)=0$; i.e., when $x=-1,3$.

We have found that f has two critical numbers, $x=-1,3$, and at $x=1$ something important might also happen. These three numbers divide the real number line into 4 subintervals:

$$(-\infty,-1),\quad (-1,1),\quad (1,3)\quad \text{and}\quad (3,\infty). \tag{3.3.5}$$

Pick a number p from each subinterval and test the sign of f' at p to determine whether f is increasing or decreasing on that interval. Again, we do well to avoid complicated computations; notice that the denominator of f' is *always* positive so we can ignore it during our work.

Interval 1, $(-\infty,-1)$: Choosing a very small number (i.e., a negative number with a large magnitude) p returns p^2-2p-3 in the numerator of f'; that will be positive. Hence f is increasing on $(-\infty,-1)$.

Interval 2, $(-1,1)$: Choosing 0 seems simple: $f'(0)=-3<0$. We conclude f is decreasing on $(-1,1)$.

Interval 3, $(1,3)$: Choosing 2 seems simple: $f'(2)=-3<0$. Again, f is decreasing.

Interval 4, $(3,\infty)$: Choosing an very large number p from this subinterval will give a positive numerator and (of course) a positive denominator. So f is increasing on $(3,\infty)$.

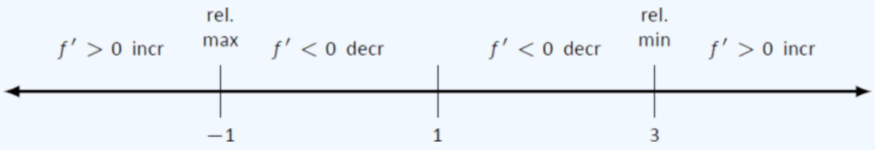


***Figure** 3.3.5**:** Number line for f in Example 3.3.2*

In summary, f is increasing on the set $(-\infty,-1)\cup(3,\infty)$ and is decreasing on the set $(-1,1)\cup(1,3)$. Since at $x=-1$, the sign of f'\ switched from positive to negative, Theorem 3.3.2 states that $f(-1)$ is a relative maximum of f. At $x=3$, the sign of f'\ switched from negative to positive, meaning $f(3)$ is a relative minimum. At $x=1$, f is not defined, so there is no relative extrema at $x=1$.

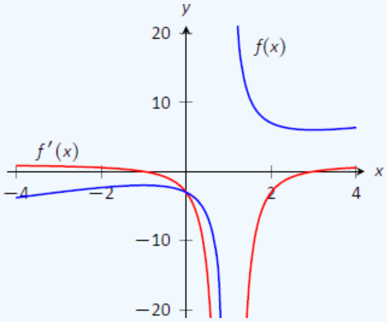


Figure $3.3.6$**:** *A graph of $f(x)$ in Example $3.3.2$, showing where f is increasing and decreasing.*

This is summarized in the number line shown in Figure $3.3.3$. Also, Figure $3.3.4$ shows a graph of f, confirming our calculations. This figure also shows f', again demonstrating that f is increasing when $f'>0$ and decreasing when $f'<0$.

One is often tempted to think that functions always alternate "increasing, decreasing, increasing, decreasing,..." around critical values. Our previous example demonstrated that this is not always the case. While $x=1$ was not technically a critical value, it was an important value we needed to consider. We found that f was decreasing on "both sides of $x=1$."

We examine one more example.

Example $3.3.3$: Using the First Derivative Test

Find the intervals on which $f(x)=x^{8/3}-4x^{2/3}$ is increasing and decreasing and identify the relative extrema.

Solution

We again start with taking derivatives. Since we know we want to solve $f'(x)=0$, we will do some algebra after taking derivatives.

$$f(x)=x^{\frac{8}{3}}-4x^{\frac{2}{3}} \tag{3.3.6}$$

$$f'(x)=\frac{8}{3}x^{\frac{5}{3}}-\frac{8}{3}x^{-\frac{1}{3}} \tag{3.3.7}$$

$$=\frac{8}{3}x^{-\frac{1}{3}}\left(x^{\frac{6}{3}}-1\right) \tag{3.3.8}$$

$$=\frac{8}{3}x^{-\frac{1}{3}}(x^2-1) \tag{3.3.9}$$

$$=\frac{8}{3}x^{-\frac{1}{3}}(x-1)(x+1). \tag{3.3.10}$$

This derivation of f' shows that $f'(x)=0$ when $x=\pm 1$ and f' is not defined when $x=0$. Thus we have 3 critical values, breaking the number line into 4 subintervals as shown in Figure $3.3.5$.

Interval 1, $(\infty,-1)$: We choose $p=-2$; we can easily verify that $f'(-2)<0$. So f is decreasing on $(-\infty,-1)$.

Interval 2, $(-1,0)$: Choose $p=-1/2$. Once more we practice finding the sign of $f'(p)$ without computing an actual value. We have \$$f'(p)=(8/3)p^{-1/3}(p-1)(p+1)$; find the sign of each of the three terms.

$$f'(p)=\frac{8}{3}\cdot\underbrace{p^{-\frac{1}{3}}}_{<0}\cdot\underbrace{(p-1)}_{<0}\underbrace{(p+1)}_{>0}. \tag{3.3.11}$$

We have a "negative × negative × positive" giving a positive number; f is increasing on $(-1,0)$.

Interval 3, $(0,1)$: We do a similar sign analysis as before, using p in $(0,1)$.

$$f'(p)=\frac{8}{3}\cdot\underbrace{p^{-\frac{1}{3}}}_{>0}\cdot\underbrace{(p-1)}_{<0}\underbrace{(p+1)}_{>0}. \tag{3.3.12}$$

We have 2 positive factors and one negative factor; $f'(p) < 0$ and so f is decreasing on $(0, 1)$.

Interval 4, $(1, \infty)$: Similar work to that done for the other three intervals shows that $f'(x) > 0$ on $(1, \infty)$, so f is increasing on this interval.

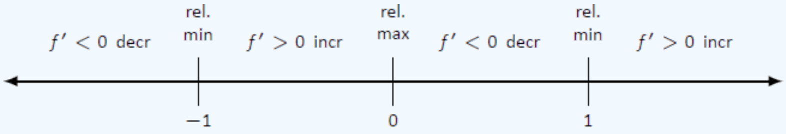


Figure **3.3.7*:*** *Number line for f in Example 3.3.3*

We conclude by stating that f is increasing on $(-1, 0) \cup (1, \infty)$ and decreasing on $(-\infty, -1) \cup (0, 1)$. The sign of f' changes from negative to positive around $x = -1$ and $x = 1$, meaning by Theorem 3.3.2 that $f(-1)$ and $f(1)$ are relative minima of f. As the sign of f' changes from positive to negative at $x = 0$, we have a relative maximum at $f(0)$. Figure 3.3.8 shows a graph of f, confirming our result. We also graph f', highlighting once more that f is increasing when $f' > 0$ and is decreasing when $f' < 0$.

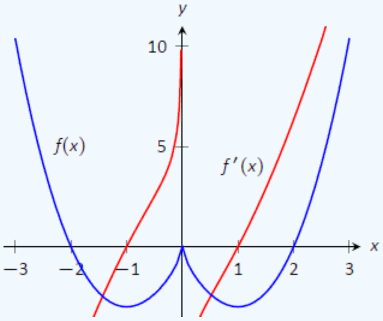


Figure 3.3.8: *A graph of $f(x)$ in Example 3.3.3, showing where f is increasing and decreasing.*

We have seen how the first derivative of a function helps determine when the function is going "up" or "down." In the next section, we will see how the second derivative helps determine how the graph of a function curves.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/
- Integrated by Justin Marshall.

This page titled 3.3: Increasing and Decreasing Functions is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

3.4: Concavity and the Second Derivative

Our study of "nice" functions continues. The previous section showed how the first derivative of a function, f', can relay important information about f. We now apply the same technique to f' itself, and learn what this tells us about f. The key to studying f' is to consider its derivative, namely f'', which is the second derivative of f. When $f''>0$, f' is increasing. When $f''<0$, f' is decreasing. f' has relative maxima and minima where $f''=0$ or is undefined. This section explores how knowing information about f'' gives information about f.

Concavity

We begin with a definition, then explore its meaning.

Definition Concave Up and Concave Down

Let f be differentiable on an interval I. The graph of f is **concave up** on I if f' is increasing. The graph of f is **concave down** on I if f' is decreasing. If f' is constant then the graph of f is said to have **no concavity.**

Note: We often state that "f is concave up" instead of "the graph of f is concave up" for simplicity.

The graph of a function f is *concave up* when f' is *increasing.* That means as one looks at a concave up graph from left to right, the slopes of the tangent lines will be increasing. Consider Figure $3.4.1$, where a concave up graph is shown along with some tangent lines. Notice how the tangent line on the left is steep, downward, corresponding to a small value of f'. On the right, the tangent line is steep, upward, corresponding to a large value of f'.

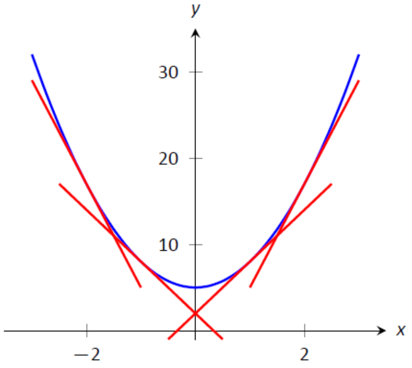


***Figure** $3.4.1$**:** A function f with a concave up graph. Notice how the slopes of the tangent lines, when looking from left to right, are increasing.*

If a function is decreasing and concave up, then its rate of decrease is slowing; it is "leveling off." If the function is increasing and concave up, then the *rate* of increase is increasing. The function is increasing at a faster and faster rate.

Now consider a function which is concave down. We essentially repeat the above paragraphs with slight variation.

The graph of a function f is *concave down* when f' is *decreasing*. That means as one looks at a concave down graph from left to right, the slopes of the tangent lines will be decreasing. Consider Figure $3.4.2$, where a concave down graph is shown along with some tangent lines. Notice how the tangent line on the left is steep, upward, corresponding to a large value of f'. On the right, the tangent line is steep, downward, corresponding to a small value of f'.

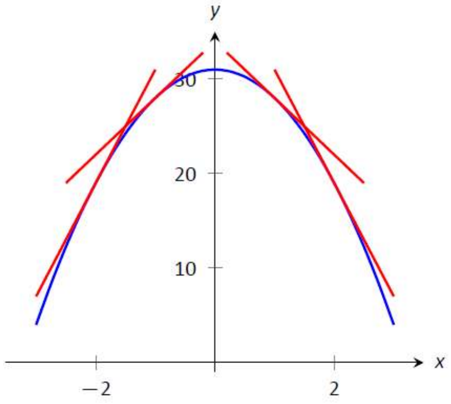


Figure $3.4.2$***:*** *A function f with a concave down graph. Notice how the slopes of the tangent lines, when looking from left to right, are decreasing.*

If a function is increasing and concave down, then its rate of increase is slowing; it is "leveling off." If the function is decreasing and concave down, then the *rate* of decrease is decreasing. The function is decreasing at a faster and faster rate.

Note: A mnemonic for remembering what concave up/down means is: "Concave up is like a cup; concave down is like a frown." It is admittedly terrible, but it works.

Our definition of concave up and concave down is given in terms of when the first derivative is increasing or decreasing. We can apply the results of the previous section and to find intervals on which a graph is concave up or down. That is, we recognize that f' is increasing when $f'' > 0$, etc.

THeorem $3.4.1$: Test for Concavity

Let f be twice differentiable on an interval I. The graph of f is concave up if $f'' > 0$ on I, and is concave down if $f'' < 0$ on I.

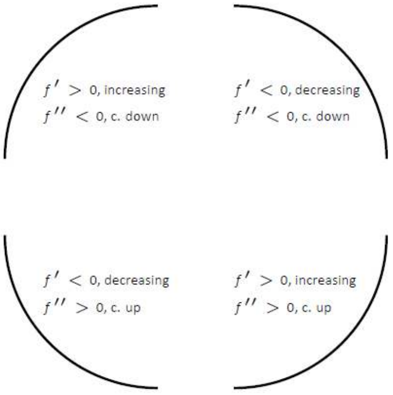


Figure $3.4.3$***:*** *Demonstrating the 4 ways that concavity interacts with increasing/decreasing, along with the relationships with the first and second derivatives.*

Note: Geometrically speaking, a function is concave up if its graph lies above its tangent lines. A function is concave down if its graph lies below its tangent lines.

If knowing where a graph is concave up/down is important, it makes sense that the places where the graph changes from one to the other is also important. This leads us to a definition.

Definition: Point of Inflection

A **point of inflection** is a point on the graph of f at which the concavity of f changes.

Figure $3.4.4$ shows a graph of a function with inflection points labeled.

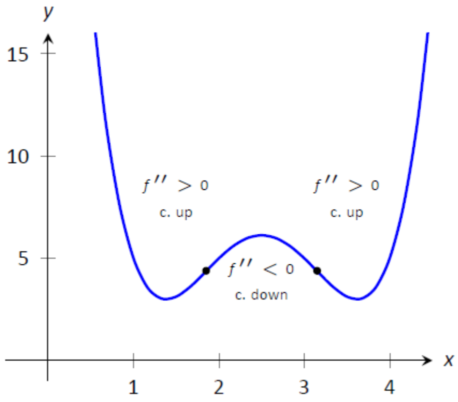


***Figure** 3.4.4: A graph of a function with its inflection points marked. The intervals where concave up/down are also indicated.*

If the concavity of f changes at a point $(c, f(c))$, then f' is changing from increasing to decreasing (or, decreasing to increasing) at $x = c$. That means that the sign of f'' is changing from positive to negative (or, negative to positive) at $x = c$. This leads to the following theorem.

THeorem 3.4.2: Points of Inflection

If $(c, f(c))$ is a point of inflection on the graph of f, then either $f'' = 0$ or f'' is not defined at c.

We have identified the concepts of concavity and points of inflection. It is now time to practice using these concepts; given a function, we should be able to find its points of inflection and identify intervals on which it is concave up or down. We do so in the following examples.

Example 3.4.1: Finding intervals of concave up/down, inflection points

Let $f(x) = x^3 - 3x + 1$. Find the inflection points of f and the intervals on which it is concave up/down.

Solution

We start by finding $f'(x) = 3x^2 - 3$ and $f''(x) = 6x$. To find the inflection points, we use Theorem 3.4.2 and find where $f''(x) = 0$ or where f'' is undefined. We find f'' is always defined, and is 0 only when $x = 0$. So the point $(0, 1)$ is the only possible point of inflection.

This possible inflection point divides the real line into two intervals, $(-\infty, 0)$ and $(0, \infty)$. We use a process similar to the one used in the previous section to determine increasing/decreasing. Pick any $c < 0$; $f''(c) < 0$ so f is concave down on $(-\infty, 0)$. Pick any $c > 0$; $f''(c) > 0$ so f is concave up on $(0, \infty)$. Since the concavity changes at $x = 0$, the point $(0, 1)$ is an inflection point.

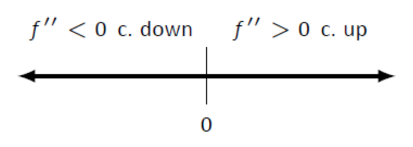


***Figure** 3.4.5: A number line determining the concavity of f in Example 3.4.1.*

The number line in Figure 3.4.5 illustrates the process of determining concavity; Figure 3.4.6 shows a graph of f and f'', confirming our results. Notice how f is concave down precisely when $f''(x) < 0$ and concave up when $f''(x) > 0$.

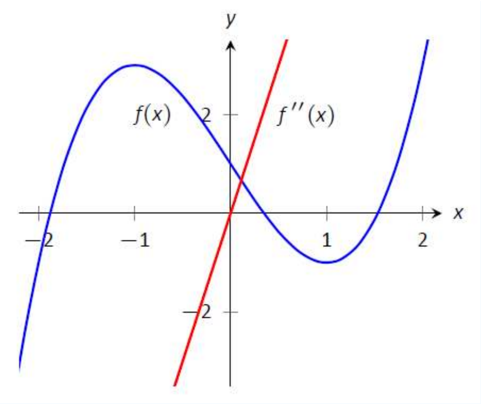


***Figure** $3.4.6$: A graph of $f(x)$ used in Example$3.4.1$*

Example $3.4.2$: Finding intervals of concave up/down, inflection points

Let $f(x) = x/(x^2 - 1)$. Find the inflection points of f and the intervals on which it is concave up/down.

Solution

We need to find f' and f''. Using the Quotient Rule and simplifying, we find

$$f'(x) = \frac{-(1+x^2)}{(x^2-1)^2} \quad \text{and} \quad f''(x) = \frac{2x(x^2+3)}{(x^2-1)^3}. \tag{3.4.1}$$

To find the possible points of inflection, we seek to find where $f''(x) = 0$ and where f'' is not defined. Solving $f''x) = 0$ reduces to solving $2x(x^2+3) = 0$; we find $x = 0$. We find that f'' is not defined when $x = \pm 1$, for then the denominator of f'' is 0. We also note that f itself is not defined at $x = \pm 1$, having a domain of $(-\infty, -1) \cup (-1, 1) \cup (1, \infty)$. Since the domain of f is the union of three intervals, it makes sense that the concavity of f could switch across intervals. We technically cannot say that f has a point of inflection at $x = \pm 1$ as they are not part of the domain, but we must still consider these x-values to be important and will include them in our number line.

The important x-values at which concavity might switch are $x = -1$, $x = 0$ and $x = 1$, which split the number line into four intervals as shown in Figure $3.4.7$. We determine the concavity on each. Keep in mind that all we are concerned with is the *sign* of f'' on the interval.

Interval 1, $(-\infty, -1)$: Select a number c in this interval with a large magnitude (for instance, $c = -100$). The denominator of $f''(x)$ will be positive. In the numerator, the $(c^2 + 3)$ will be positive and the $2c$ term will be negative. Thus the numerator is negative and $f''(c)$ is negative. We conclude f is concave down on $(-\infty, -1)$.

Interval 2, $(-1, 0)$: For any number c in this interval, the term $2c$ in the numerator will be negative, the term $(c^2 + 3)$ in the numerator will be positive, and the term $(c^2 - 1)^3$ in the denominator will be negative. Thus $f''(c) > 0$ and f is concave up on this interval.

Interval 3, $(0, 1)$: Any number c in this interval will be positive and "small." Thus the numerator is positive while the denominator is negative. Thus $f''(c) < 0$ and f is concave down on this interval.

Interval 4, $(1, \infty)$: Choose a large value for c. It is evident that $f''(c) > 0$, so we conclude that f is concave up on $(1, \infty)$.

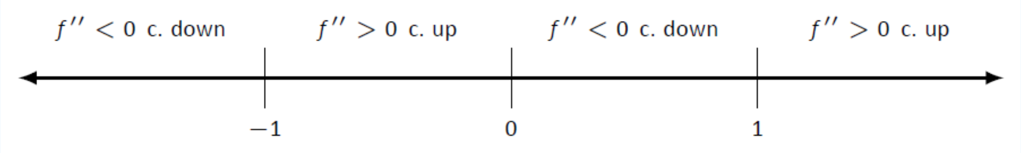


***Figure** $3.4.7$: Number line for f in Example $3.4.2$*

We conclude that f is concave up on $(-1, 0) \cup (1, \infty)$ and concave down on $(-\infty, -1) \cup (0, 1)$. There is only one point of inflection, $(0, 0)$, as f is not defined at $x = \pm 1$. Our work is confirmed by the graph of f in Figure $3.4.8$. Notice how f is concave up whenever f'' is positive, and concave down when f'' is negative.

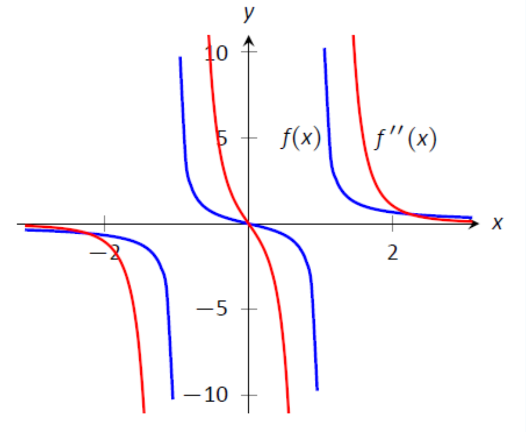


Figure $3.4.8$***:*** *A graph of* $f(x)$ *and* $f''(x)$ *in Example* $3.4.2$

Recall that relative maxima and minima of f are found at critical points of f; that is, they are found when $f'(x) = 0$ or when f' is undefined. Likewise, the relative maxima and minima of f' are found when $f''(x) = 0$ or when f'' is undefined; note that these are the inflection points of f.

What does a "relative maximum of f'" *mean*? The derivative measures the rate of change of f; maximizing f' means finding the where f is increasing the most -- where f has the steepest tangent line. A similar statement can be made for minimizing f'; it corresponds to where f has the steepest negatively--sloped tangent line.

We utilize this concept in the next example.

Example $3.4.3$: Understanding inflection points

The sales of a certain product over a three-year span are modeled by $S(t) = t^4 - 8t^2 + 20$, where t is the time in years, shown in Figure $3.4.9$. Over the first two years, sales are decreasing. Find the point at which sales are decreasing at their greatest rate.

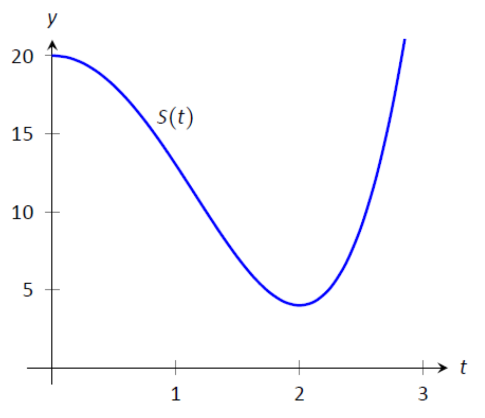


Figure $3.4.9$***:*** *A graph of* $S(t)$ *in Example* $3.4.3$*, modeling the sale of a product over time.*

We want to maximize the rate of decrease, which is to say, we want to find where S' has a minimum. To do this, we find where S'' is 0. We find $S'(t) = 4t^3 - 16t$ and $S''(t) = 12t^2 - 16$. Setting $S''(t) = 0$ and solving, we get $t = \sqrt{4/3} \approx 1.16$ (we ignore the negative value of t since it does not lie in the domain of our function S).

This is both the inflection point and the point of maximum decrease. This is the point at which things first start looking up for the company. After the inflection point, it will still take some time before sales start to increase, but at least sales are not decreasing quite as quickly as they had been.

A graph of $S(t)$ and $S'(t)$ is given in Figure $3.4.10$. When $S'(t) < 0$, sales are decreasing; note how at $t \approx 1.16$, $S'(t)$ is minimized. That is, sales are decreasing at the fastest rate at $t \approx 1.16$. On the interval of $(1.16, 2)$, S is decreasing but concave up, so the decline in sales is "leveling off."

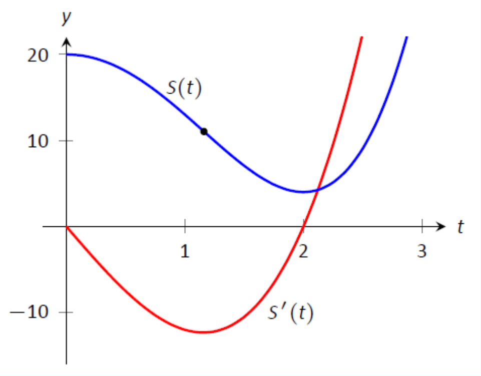


Figure $3.4.10$***:*** *A graph of* $S(t)$ *in Example* $3.4.3$ *along with* $S'(t)$.

Not every critical point corresponds to a relative extrema; $f(x)=x^3$ has a critical point at $(0,0)$ but no relative maximum or minimum. Likewise, just because $f''(x)=0$ we cannot conclude concavity changes at that point. We were careful before to use terminology "*possible* point of inflection" since we needed to check to see if the concavity changed. The canonical example of $f''(x)=0$ *without* concavity changing is $f(x)=x^4$. At $x=0$, $f''(x)=0$ but f is always concave up, as shown in Figure $3.4.11$.

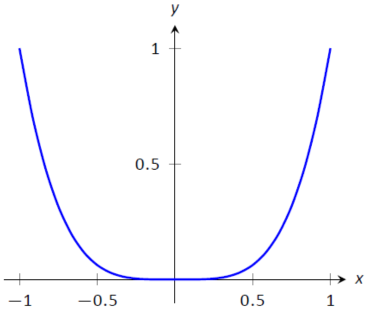


Figure $3.4.11$***:*** *A graph of* $f(x)=x^4$*. Clearly* f *is always concave up, despite the fact that* $f''(x)=0$ *when* $x=0$*. It this example, the possible point of inflection* $(0,0)$ *is not a point of inflection.*

Note: A note about concavity:

In general, concavity can change only where either the second derivative is 0, where there is a vertical asymptote, or (rare in practice) where the second derivative is undefined. But concavity doesn't \emph{have} to change at these places. For instance, if $f(x)=x^4$, then $f''(0)=0$, but there is no change of concavity at 0 and also no inflection point there. Moreover, if $f(x)=1/x^2$, then f has a vertical asymptote at 0, but there is no change in concavity at 0.

The Second Derivative Test

The first derivative of a function gave us a test to find if a critical value corresponded to a relative maximum, minimum, or neither. The second derivative gives us another way to test if a critical point is a local maximum or minimum. The following theorem officially states something that is intuitive: if a critical value occurs in a region where a function f is concave up, then that critical value must correspond to a relative minimum of f, etc. See Figure $3.4.12$ for a visualization of this.

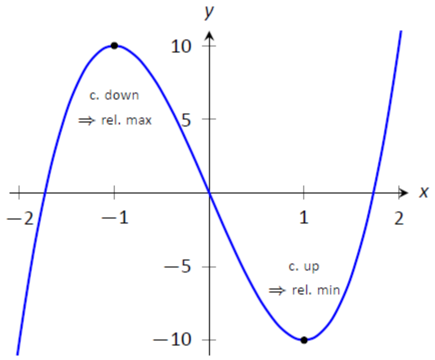


***Figure** $3.4.12$: Demonstrating the fact that relative maxima occur when the graph is concave down and relatve minima occur when the graph is concave up.*

THeorem $3.4.3$: The Second Derivative Test

Let c be a critical value of f where $f''(c)$ is defined.

1. If $f''(c) > 0$, then f has a local minimum at $(c, f(c))$.
2. If $f''(c) < 0$, then f has a local maximum at $(c, f(c))$.

The Second Derivative Test relates to the First Derivative Test in the following way. If $f''(c) > 0$, then the graph is concave up at a critical point c and f' itself is growing. Since $f'(c) = 0$ and f' is growing at c, then it must go from negative to positive at c. This means the function goes from decreasing to increasing, indicating a local minimum at c.

Example $3.4.4$: Using the Second Derivative Test

Let $f(x) = 100/x + x$. Find the critical points of f and use the Second Derivative Test to label them as relative maxima or minima.

Solution

We find $f'(x) = -100/x^2 + 1$ and $f''(x) = 200/x^3$. We set $f'(x) = 0$ and solve for x to find the critical values (note that f'\ is not defined at $x = 0$, but neither is f so this is not a critical value.) We find the critical values are $x = \pm 10$. Evaluating f'' at $x = 10$ gives $0.1 > 0$, so there is a local minimum at $x = 10$. Evaluating $f''(-10) = -0.1 < 0$, determining a relative maximum at $x = -10$. These results are confirmed in Figure $3.4.13$.

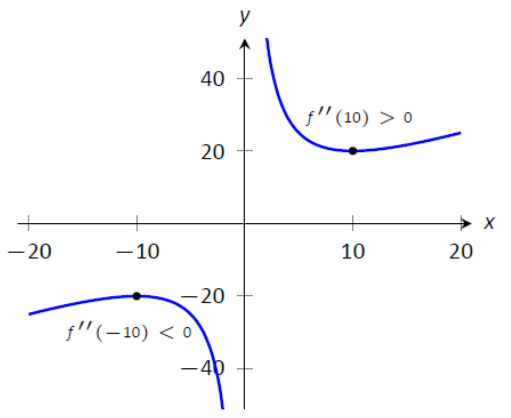


***Figure** $3.4.13$: A graph of $f(x)$ in Example $3.4.4$. The second derivative is evaluated at each critical point. When the graph is concave up, the critical point represents a local minimum; when the graph is concave down, the critical point represents a local maximum.*

We have been learning how the first and second derivatives of a function relate information about the graph of that function. We have found intervals of increasing and decreasing, intervals where the graph is concave up and down, along with the locations of relative extrema and inflection points. In Chapter 1 we saw how limits explained asymptotic behavior. In the next section we combine all of this information to produce accurate sketches of functions.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/
- Integrated by Justin Marshall.

This page titled 3.4: Concavity and the Second Derivative is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

3.5: Curve Sketching

We have been learning how we can understand the behavior of a function based on its first and second derivatives. While we have been treating the properties of a function separately (increasing and decreasing, concave up and concave down, etc.), we combine them here to produce an accurate graph of the function without plotting lots of extraneous points.

Why bother? Graphing utilities are very accessible, whether on a computer, a hand--held calculator, or a smartphone. These resources are usually very fast and accurate. We will see that our method is not particularly fast -- it will require time (but it is not *hard*). So again: why bother?

We are attempting to understand the behavior of a function f based on the information given by its derivatives. While all of a function's derivatives relay information about it, it turns out that "most" of the behavior we care about is explained by f' and f''. Understanding the interactions between the graph of f and $f'$and f' is important. To gain this understanding, one might argue that all that is needed is to look at lots of graphs. This is true to a point, but is somewhat similar to stating that one understands how an engine works after looking only at pictures. It is true that the basic ideas will be conveyed, but "hands--on" access increases understanding.

The following Key Idea summarizes what we have learned so far that is applicable to sketching graphs of functions and gives a framework for putting that information together. It is followed by several examples.

Key Idea 4: Curve Sketching

To produce an accurate sketch a given function f, consider the following steps.

1. Find the domain of f. Generally, we assume that the domain is the entire real line then find restrictions, such as where a denominator is 0 or where negatives appear under the radical.
2. Find the critical values of f.
3. Find the possible points of inflection of f.
4. Find the location of any vertical asymptotes of f (usually done in conjunction with item 1 above).
5. Consider the limits $\lim\limits_{x\to-\infty} f(x)$ and $\lim\limits_{x\to\infty} f(x)$ to determine the end behavior of the function.
6. Create a number line that includes all critical points, possible points of inflection, and locations of vertical asymptotes. For each interval created, determine whether f is increasing or decreasing, concave up or down.
7. Evaluate f at each critical point and possible point of inflection. Plot these points on a set of axes. Connect these points with curves exhibiting the proper concavity. Sketch asymptotes and x and y intercepts where applicable.

Example $3.5.1$: curve sketching

Use Key Idea 4 to sketch $f(x)=3x^3-10x^2+7x+5$.

Solution

1. The domain of f is the entire real line; there are no values x for which $f(x)$ is not defined.
2. Find the critical values of f. We compute $f'(x)=9x^2-20x+7$. Use the Quadratic Formula to find the roots of f':

$$x=\frac{20\pm\sqrt{(-20)^2-4(9)(7)}}{2(9)}=\frac{1}{9}\left(10\pm\sqrt{37}\right)\Rightarrow x\approx 0.435, 1.787. \tag{3.5.1}$$

3. Find the possible points of inflection of f. Compute $f''(x)=18x-20$. We have

$$f'p(x)=0\Rightarrow x=10/9\approx 1.111. \tag{3.5.2}$$

4. There are no vertical asymptotes.
5. We determine the end behavior using limits as x approaches ±infinity.

$$\lim_{x\to-\infty} f(x)=-\infty \qquad \lim_{x\to\infty} f(x)=\infty. \tag{3.5.3}$$

   We do not have any horizontal asymptotes.
6. We place the values $x=(10\pm\sqrt{37})/9$ and $x=10/9$ on a number line, as shown in Figure $3.5.1$. We mark each subinterval as increasing or decreasing, concave up or down, using the techniques used in Sections 3.3 and 3.4.

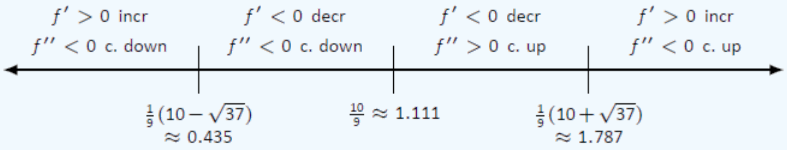


Figure *$3.5.1$**:** Number line for f in Example $3.5.1$.*

7. We plot the appropriate points on axes as shown in Figure $3.5.2a$ and connect the points with straight lines. In Figure $3.5.2b$ we adjust these lines to demonstrate the proper concavity. Our curve crosses the y axis at $y = 5$ and crosses the x axis near $x = -0.424$. In Figure $3.5.2c$ we show a graph of f drawn with a computer program, verifying the accuracy of our sketch.

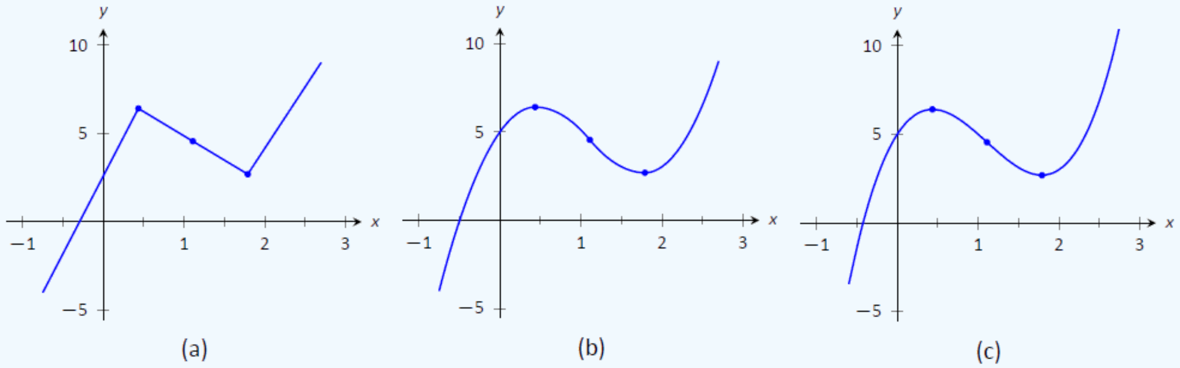


Figure *$3.5.2$**:** Sketching f in Example $3.5.1$.*

Example $3.5.2$: Curve sketching

Sketch $f(x) = \dfrac{x^2 - x - 2}{x^2 - x - 6}$.

Solution

We again follow the steps outlined in Key Idea 4.

1. In determining the domain, we assume it is all real numbers and looks for restrictions. We find that at $x = -2$ and $x = 3$, $f(x)$ is not defined. So the domain of f is $D = \{\text{real numbers } x \mid x \neq -2, 3\}$.
2. To find the critical values of f, we first find $f'(x)$. Using the Quotient Rule, we find

$$f'(x) = \frac{-8x + 4}{(x^2 + x - 6)^2} = \frac{-8x + 4}{(x-3)^2(x+2)^2}. \tag{3.5.4}$$

   $f'(x) = 0$ when $x = 1/2$, and f' is undefined when $x = -2, 3$. Since f' is undefined only when f is, these are not critical values. The only critical value is $x = 1/2$.
3. To find the possible points of inflection, we find $f''(x)$, again employing the Quotient Rule:

$$f''(x) = \frac{24x^2 - 24x + 56}{(x-3)^3(x+2)^3}. \tag{3.5.5}$$

   We find that $f''(x)$ is never 0 (setting the numerator equal to 0 and solving for x, we find the only roots to this quadratic are imaginary) and $f'$is undefined when $x = -2, 3$. Thus concavity will possibly only change at $x = -2$ and $x = 3$.
4. The vertical asymptotes of f are at $x = -2$ and $x = 3$, the places where f is undefined.
5. There is a horizontal asymptote of $y = 1$, as $\lim_{x \to -\infty} f(x) = 1$ and $\lim_{x \to \infty} f(x) = 1$.
6. We place the values $x = 1/2$, $x = -2$ and $x = 3$ on a number line as shown in Figure $3.5.3$. We mark in each interval whether f is increasing or decreasing, concave up or down. We see that f has a relative maximum at $x = 1/2$; concavity changes only at the vertical asymptotes.

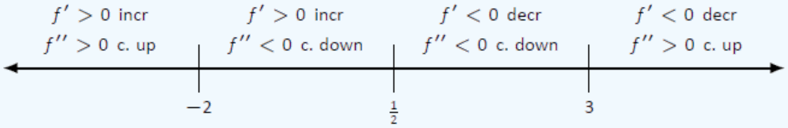
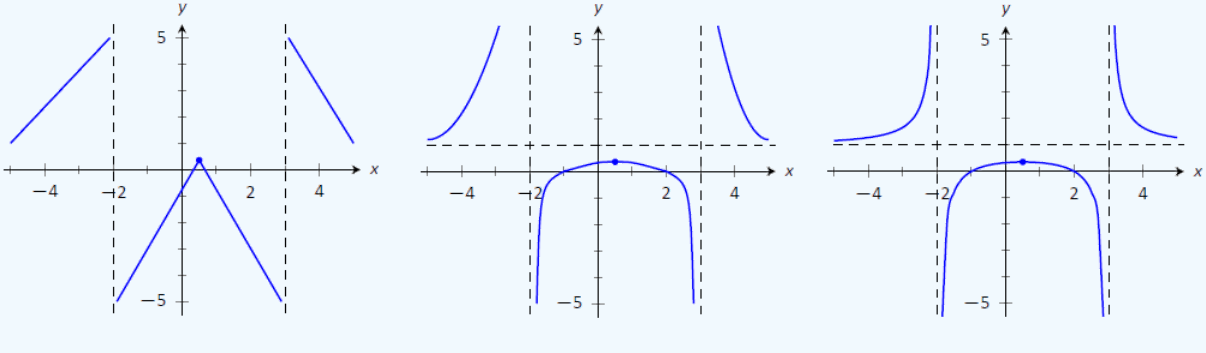

$f' > 0$ incr
$f'' > 0$ c. up
$f' > 0$ incr
$f'' < 0$ c. down
$f' < 0$ decr
$f'' < 0$ c. down
$f' < 0$ decr
$f'' > 0$ c. up
-2
$\frac{1}{2}$
3

***FIgure** $3.5.3$**:** Number line for f in Example $3.5.2$*

7. In Figure $3.5.4a$, we plot the points from the number line on a set of axes and connect the points with straight lines to get a general idea of what the function looks like (these lines effectively only convey increasing/decreasing information). In Figure $3.5.4b$, we adjust the graph with the appropriate concavity. We also show f crossing the x axis at $x = -1$ and $x = 2$.



(a) (b) (c)

***Figure** $3.5.4$**:** Sketching f in Example $3.5.2$.*

Figure $3.5.4c$ shows a computer generated graph of f, which verifies the accuracy of our sketch.

Example $3.5.3$: Curve sketching

Sketch $f(x) = \frac{5(x-2)(x+1)}{x^2+2x+4}$.

Solution

We again follow Key Idea 4

1. We assume that the domain of f is all real numbers and consider restrictions. The only restrictions come when the denominator is 0, but this never occurs. Therefore the domain of f is all real numbers, $\mathbb{R}$.
2. We find the critical values of f by setting $f'(x) = 0$ and solving for x. We find

$$f'(x) = \frac{15x(x+4)}{(x^2+2x+4)^2} \quad \Rightarrow \quad f'(x) = 0 \text{ when } \ x = -4, 0. \tag{3.5.6}$$

3. We find the possible points of inflection by solving $f''(x) = 0$ for x. We find

$$f'p(x) = -\frac{30x^3+180x^2-240}{(x^2+2x+4)^3}. \tag{3.5.7}$$

   The cubic in the numerator does not factor very "nicely." We instead approximate the roots at $x = -5.759$, $x = -1.305$ and $x = 1.064$.
4. There are no vertical asymptotes.
5. We have a horizontal asymptote of $y = 5$, as $\lim_{x\to-\infty} f(x) = \lim_{x\to\infty} f(x) = 5$.
6. We place the critical points and possible points on a number line as shown in Figure $3.5.5$ and mark each interval as increasing/decreasing, concave up/down appropriately.

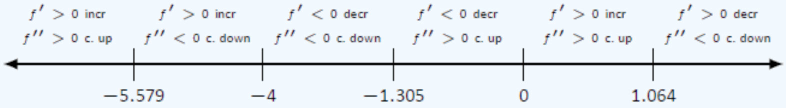

$f' > 0$ incr
$f'' > 0$ c. up
$f' > 0$ incr
$f'' < 0$ c. down
$f' < 0$ decr
$f'' < 0$ c. down
$f' < 0$ decr
$f'' > 0$ c. up
$f' > 0$ incr
$f'' > 0$ c. up
$f' > 0$ decr
$f'' < 0$ c. down
-5.579
-4
-1.305
0
1.064

***Figure** $3.5.5$**:** Number line for f in Example $3.5.3$.*

7. In Figure $3.5.6a$ we plot the significant points from the number line as well as the two roots of f, $x=-1$ and $x=2$, and connect the points with straight lines to get a general impression about the graph. In Figure $3.5.6b$, we add concavity. Figure $3.5.6c$ shows a computer generated graph of f, affirming our results.

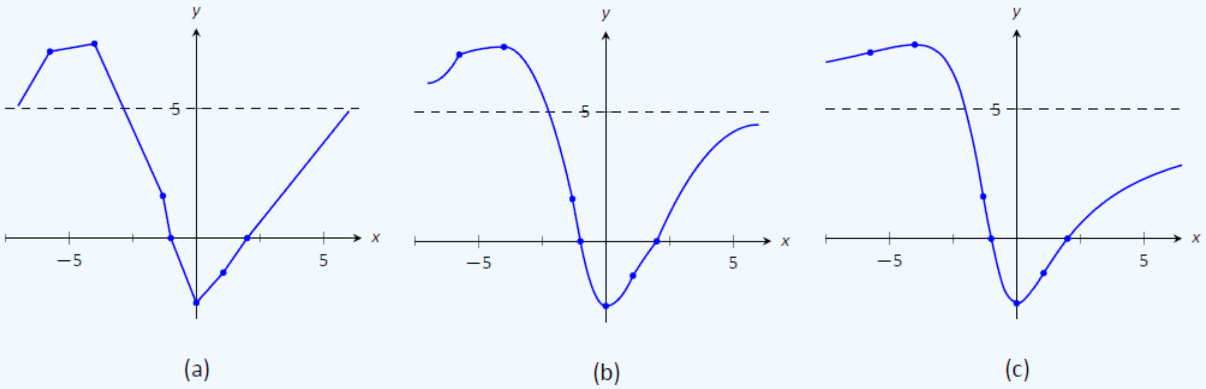


***Figure** $3.5.6$**:** Sketching f in Example $3.5.3$.*

In each of our examples, we found a few, significant points on the graph of f that corresponded to changes in increasing/decreasing or concavity. We connected these points with straight lines, then adjusted for concavity, and finished by showing a very accurate, computer generated graph.

Why are computer graphics so good? It is not because computers are "smarter" than we are. Rather, it is largely because computers are much faster at computing than we are. In general, computers graph functions much like most students do when first learning to draw graphs: they plot equally spaced points, then connect the dots using lines. By using lots of points, the connecting lines are short and the graph looks smooth.

This does a fine job of graphing in most cases (in fact, this is the method used for many graphs in this text). However, in regions where the graph is very "curvy," this can generate noticeable sharp edges on the graph unless a large number of points are used. High quality computer algebra systems, such as *Mathematica*, use special algorithms to plot lots of points only where the graph is "curvy."

In Figure $3.5.7$, a graph of $y=\sin x$ is given, generated by *Mathematica*. The small points represent each of the places *Mathematica* sampled the function. Notice how at the "bends" of $\sin x$, lots of points are used; where $\sin x$ is relatively straight, fewer points are used. (Many points are also used at the endpoints to ensure the "end behavior" is accurate.)

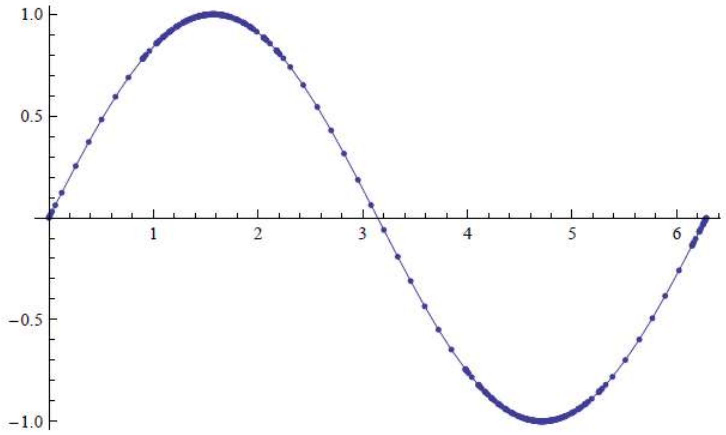


***Figure** $3.5.7$**:** A graph of $y=\sin x$ generated by Mathematica.*

How does *Mathematica* know where the graph is "curvy"? Calculus. When we study *curvature* in a later chapter, we will see how the first and second derivatives of a function work together to provide a measurement of "curviness." *Mathematica* employs algorithms to determine regions of "high curvature"' and plots extra points there.

Again, the goal of this section is not "How to graph a function when there is no computer to help." Rather, the goal is "Understand that the shape of the graph of a function is largely determined by understanding the behavior of the function at a few key places." In Example $3.5.3$, we were able to accurately sketch a complicated graph using only 5 points and knowledge of asymptotes!

There are many applications of our understanding of derivatives beyond curve sketching. The next chapter explores some of these applications, demonstrating just a few kinds of problems that can be solved with a basic knowledge of differentiation.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/
- Integrated by Justin Marshall.

This page titled 3.5: Curve Sketching is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

3.E: Applications of the Graphical Behavior of Functions(Exercises)

3.1: Extreme Values

Terms and Concepts

1. Describe what an "extreme value" of a function is in your own words.

2. Sketch the graph of a function f on (-1,1) that has both a maximum and minimum value.

3. Describe the difference between absolute and relative maxima in your own words.

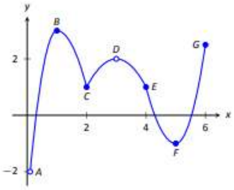
4. Sketch the graph of a function f where f has a relative maximum at $x=1$ and $f'(1)$ is undefined.

5. T/F: If c is a critical value of a function f, then f has either a relative maximum or relative minimum at $x=c$.

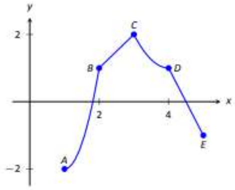
Problems

In Exercises 6-7, identify each of the marked points as being an absolute maximum and minimum, a relative maximum or minimum, or none of the above.

6.

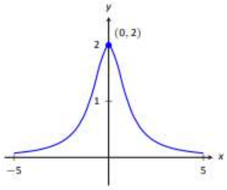


7.

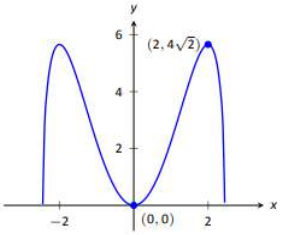


In Exercises 8-14, evaluate $f'(x)$ at the points indicated in the graph.

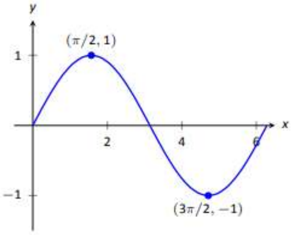
8. $f(x)=\frac{2}{x^2+1}$



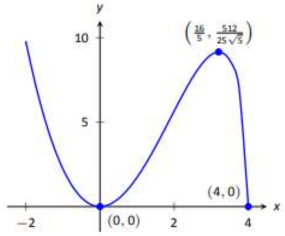
9. $f(x)=x^2\sqrt{6-x^2}$



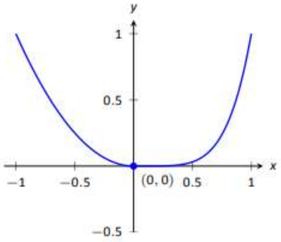
10. $f(x) = \sin x$



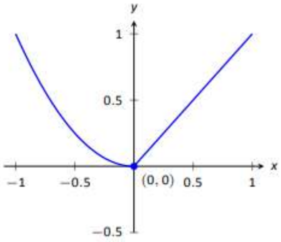
11. $f(x) = x^2\sqrt{4-x}$



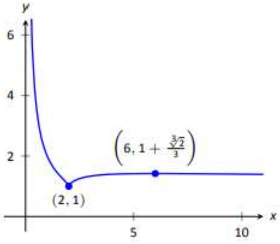
12. $f(x) = \begin{cases} x^2 & x \le 0 \\ x^5 & x > 0 \end{cases}$



13. $f(x) = \begin{cases} x^2 & x \le 0 \\ x & x > 0 \end{cases}$



14. $f(x) = \frac{(x-2)^{2/3}}{x}$



In Exercises 15-24, find the extreme values of the function on the given interval.

15. $f(x) = x^2 + x + 4$ on $[-1, 2]$.

16. $f(x) = x^3 - \frac{9}{2}x^2 - 30x + 3$ on $[0, 6]$.

17. $f(x) = 3\sin x$ on $[\pi/4, 2\pi/3]$.

18. $f(x)=x^2\sqrt{4-x^2}$ on $[-2,2]$.

19. $f(x)=x+\frac{3}{x}$ on $[1,5]$.

20. $f(x)=\frac{x^2}{x^2+5}$ on $[-3,5]$.

21. $f(x)=e^x\cos x$ on $[0,\pi]$.

22. $f(x)=e^x\sin x$ on $[0,\pi]$.

23. $f(x)=\frac{\ln x}{x}$ on $[1,4]$.

24. $f(x)=x^{2/3}-x$ on $[0,2]$.

Review

25. Find $\frac{dy}{dx}$, where $x^2y-y^2x=1$.

26. Find the equation of the line tangent to the graph of $x^2+y^2+xy=7$ at the point $(1,2)$.

27. Let $f(x)=x^3+x$. Evaluate $\lim\limits_{s\to 0}\frac{f(x+s)-f(x)}{s}$.

3.2: The Mean Value Theorem

Terms and Concepts

1. Explain in your own words what the Mean Value Theorem states.

2. Explain in your own words what Rolle's Theorem states.

Problems

In Exercises 3-10, a function $f(x)$ **and interval** [a,b] **are given. Check if Rolle's Theorem can be applied to** f **on** [a,b]**; if so, find** c **in** [a,b] **such that** $f'(c)=0$.

3. $f(x)=6$ on $[-1,1]$.

4. $f(x)=6x$ on $[-1,1]$.

5. $f(x)=x^2+x-6$ on $[-3,2]$.

6. $f(x)=x^2+x-2$ on $[-3,2]$.

7. $f(x)=x^2+x$ on $[-2,2]$.

8. $f(x)=\sin x$ on $[\pi/6,5\pi/6]$.

9. $f(x)=\cos x$ on $[0,\pi]$.

10. $f(x)=\frac{1}{x^2-2x+1}$ on $[0,2]$.

In Exercises 11-20, a function $f(x)$ **and interval** [a,b] **are given. Check if the Mean Value Theorem can be applied to** f **on** [a,b]**; if so, find a value** c **in** [a,b] **guaranteed by the Mean Value Theorem.**

11. $f(x)=x^2+3x-1$ on $[-2,2]$.

12. $f(x)=5x^2-6x+8$ on $[0,5]$.

13. $f(x)=\sqrt{9-x^2}$ on $[0,3]$.

14. $f(x)=\sqrt{25-x}$ on $[0,9]$.

16. $f(x)=\ln x$ on $[1,5]$.

17. $f(x)=\tan x$ on $[\pi/4,\pi/4]$.

18. $f(x)=x^3-2x^2+x+1$ on $[-2,2]$.

19. $f(x)=2x^3-5x^2+6x+1$ on $[-5,2]$.

20. $f(x)=\sin^{-1}x$ on $[-1,1]$.

Review

21. Find the extreme values of $f(x)=x^2-3x+9$ on $[-2,5]$.

22. Describe the critical points of $f(x)=\cos x$.

23. Describe the critical points of $f(x)=\tan x$.

3.3: Increasing and Decreasing Functions

Terms and Concepts

1. In your own words describe what it means for a function to be increasing.

2. What does a decreasing function “look like”?

3. Sketch a graph of a function on [0,2] that is increasing but not strictly increasing.

4. Give an example of a function describing a situation where it is “bad” to be increasing and “good” to be decreasing.

5. A function f has derivative $f'(x)=(\sin x+2)e^{x^2+1}$, where $f'(x)>1$ for all x. Is f increasing, decreasing, or can we not tell from the given information?

Problems

In Exercises 6-13, a function $f(x)$ is given.
(a) Compute $f'(x)$.
(b) Graph f and f' on the same axes (using technology is permitted) and verify Theorem 29.

6. $f(x)=3x+4$

7. $f(x)=x^2-3x+5$

8. $f(x)=\cos x$

9. $f(x)=\tan x$

10. $f(x)=x^3-5x^2+7x-1$

11. $f(x)=2x^3-x^2+x-1$

12. $f(x)=x^4-5x^2+4$

13. $f(x)=\frac{1}{x^2+1}$

In Exercises 14-23, a function $f(x)$ is given.
(a) Give the domain of f.
(b) Find the critical numbers of f.
(c) Create a number line to determine the intervals on which f is increasing and decreasing.
(d) Use the First Derivative Test to determine whether each critical point is a relative maximum, minimum, or neither.

14. $f(x)=x^2+2x-3$

15. $f(x)=x^3+3x^2+3$

16. $f(x)=2x^3+x^2+3$

17. $f(x)=x^3-3x^2+3x-1$

18. $f(x)=\frac{1}{x^2-2x+2}$

19. $f(x)=\frac{x^2-4}{x^2-1}$

20. $f(x)=\frac{x}{x^2-2x-8}$

21. $f(x)=\frac{(x-2)^{2/3}}{x}$

22. $f(x)=\sin x\cos x$ on $(-\pi,\pi)$

23. $f(x)=5x^2-5x$

Review

24. Consider $f(x)=x^2-3x+5$ on [-1,2]; find c guaranteed by the Mean Value Theorem.

25. Consider $f(x)=\sin x$ on $[-\pi/2,\pi/2]$; find \(c) guaranteed by the Mean Value Theorem.

3.4: Concavity and the Second Derivative

Terms and Concepts

1. Sketch a graph of a function $f(x)$ that is concave up on (0,1) and is concave down on (1,2).

2. Sketch a graph of a function $f(x)$ that is:
(a) Increasing, concave up on (0,1),
(b) increasing, concave down on (1,2),
(c) decreasing, concave down on (2,3) and
(d) increasing, concave down on (3,4).

3. Is it possible for a function to be increasing and concave down on $(0,\infty)$ with a horizontal asymptote of $y=1$? If so, give a sketch of such a function.

4. Is it possible for a function to be increasing and concave upon $(0,\infty)$ with a horizontal asymptote of $y=1$? If so, give a sketch of such a function.

Problems

In Exercises 5-15, a function $f(x)$ is given.
(a) Compute $f''(x)$.
(b) Graph f and f'' on the same axes (using technology is permitted) and verify Theorem 31.

5. $f(x)=-7x+3$

6. $f(x)=-4x^2+3x-8$

7. $f(x)=4x^2+3x-8$

8. $f(x)=x^3-3x^2+x-1$

9. $f(x)=-x^3+x^2-2x+5$

10. $f(x)=\cos x$

11. $f(x)=\sin x$

12. $f(x)=\tan x$

13. $f(x)=\frac{1}{x^2+1}$

14. $f(x)=\frac{1}{x}$

15. $f(x)=\frac{1}{x^2}$

In Exercises 16-28, a function $f(x)$ is given.
(a) Find possible points of inflection of f
(b) Create a number line to determine the intervals on which f is concave up or concave down.

16. $f(x)=x^2-2x+1$

17. $f(x)=-x^2-5x+7$

18. $f(x)=x^3-x+1$

19. $f(x)=2x^3-3x^2+9x+5$

20. $f(x)=\frac{x^4}{4}+\frac{x^3}{3}-2x+3$

21. $f(x)=-3x^4+8x^3+6x^2-24x+2$

22. $f(x)=x^4-4x^3+6x^2-4x+1$

23. $f(x)=\frac{1}{x^2+1}$

24. $f(x)=\frac{x}{x^2-1}$

25. $f(x)=\sin x+\cos x$ on $(-\pi,\pi)$

26. $f(x)=x^2e^x$

27. $f(x)=x^2\ln x$

28. $f(x)=e^{-x^2}$

In Exercises 29-41, a function $f(x)$ is given. Find the critical points of f and use the Second Derivative Test, when possible, to determine the relative extrema. (Note: these are the same functions as in exercises 16-28.)

29. $f(x)=x^2-2x+1$

30. $f(x)=-x^2-5x+7$

31. $f(x)=x^3-x+1$

32. $f(x)=2x^3-3x^2+9x+5$

33. $f(x)=\frac{x^4}{4}+\frac{x^3}{3}-2x+3$

34. $f(x)=-3x^4+8x^3+6x^2-24x+2$

35. $f(x)=x^4-4x^3+6x^2-4x+1$

36. $f(x)=\frac{1}{x^2+1}$

37. $f(x)=\frac{x}{x^2-1}$

38. $f(x)=\sin x+\cos x$ on $(-\pi,\pi)$

39. $f(x)=x^2e^x$

40. $f(x)=x^2\ln x$

41. $f(x)=e^{-x^2}$

In Exercises 42-54, a function $f(x)$ is given. Find the x values where $f'(x)$ has a relative maximum or minimum. (Note: these are the same functions as in Exercises 16-28.)

42. $f(x)=x^2-2x+1$

43. $f(x)=-x^2-5x+7$

44. $f(x)=x^3-x+1$

45. $f(x)=2x^3-3x^2+9x+5$

46. $f(x)=\frac{x^4}{4}+\frac{x^3}{3}-2x+3$

47. $f(x)=-3x^4+8x^3+6x^2-24x+2$

48. $f(x)=x^4-4x^3+6x^2-4x+1$

49. $f(x)=\frac{1}{x^2+1}$

50. $f(x)=\frac{x}{x^2-1}$

51. $f(x)=\sin x+\cos x$ on $(-\pi,\pi)$

52. $f(x)=x^2e^x$

53. $f(x)=x^2\ln x$

54. $f(x)=e^{-x^2}$

3.5: Curve Sketching

Terms and Concepts

1. Why is sketching curves by hand beneficial even though technology is ubiquitous?

2. What does “ubiquitous” mean?

3. T/F: When sketching graphs of functions, it is useful to find the critical points.

4. T/F: When sketching graphs of functions, it is useful to find the possible points of inflection.

5. T/F: When sketching graphs of functions, it is useful to find the horizontal and vertical asymptotes.

Problems

In Exercises 6-11, practice using Key Idea 4 by applying the principles to the given functions with familiar graphs.

6. $f(x)=2x+4$

7. $f(x)=-x^2+1$

8. $f(x)=\sin x$

9. $f(x)=e^x$

10. $f(x)=\frac{1}{x}$

11. $f(x)=\frac{1}{x^2}$

In Exercises 12-25, sketch a graph of the given function using Key Idea 4. Show all work; check your answer with technology.

12. $f(x)=x^3-2x^2+4x+1$

13. $f(x)=-x^3+5x^2-3x+2$

14. $f(x)=x^3+3x^2+3x+1$

15. $f(x)=x^3-x^2-x+1$

16. $f(x)=(x-2)\ln(x-2)$

17. $f(x)=(x-2)^2\ln(x-2)$

18. $f(x)=\frac{x^2-4}{x^2}$

19. $f(x)=\frac{x^2-4x+3}{x^2-6x+8}$

20. $f(x)=\frac{x^2-2x+1}{x^2-6x+8}$

21. $f(x)=x\sqrt{x+1}$

22. $f(x)=x^2e^x$

23. $f(x)=\sin x\cos x \text{ on } [-\pi,\pi]$

24. $f(x)=(x-3)^{2/3}+2$

25. $f(x)=\frac{(x-1)^{2/3}}{x}$

In Exercises 26-28, a function with the parameters a and b are given. Describe the critical points and possible points of inflection of f in terms of a and b.

26. $f(x)=\frac{a}{x^2+b^2}$

27. $f(x)=\sin(ax+b)$

28. $f(x)=(x-a)(x-b)$

29. Given $x^2+y^2=1$, use implicit differentiation to find $fracdydx$ and $\frac{d^2y}{dx^2}$. Use this information to justify the sketch of the unit circle.

3.E: Applications of the Graphical Behavior of Functions(Exercises) is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

4: Applications of the Derivative

4.1: Newton's Method
4.2: Related Rates
4.3: Optimization
4.4: Differentials
4.E: Applications of Derivatives (Exercises)

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 4: Applications of the Derivative is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

4.1: Newton's Method

In Chapter 3, we learned how the first and second derivatives of a function influence its graph. In this chapter we explore other applications of the derivative.

Solving equations is one of the most important things we do in mathematics, yet we are surprisingly limited in what we can solve analytically. For instance, equations as simple as $x^5+x+1=0$ or $\cos x=x$ cannot be solved by algebraic methods in terms of familiar functions. Fortunately, there are methods that can give us *approximate* solutions to equations like these. These methods can usually give an approximation correct to as many decimal places as we like. In Section 1.5, we learned about the Bisection Method. This section focuses on another technique (which generally works faster), called Newton's Method.

Newton's Method is built around tangent lines. The main idea is that if x is sufficiently close to a root of $f(x)$, then the tangent line to the graph at $(x, f(x))$ will cross the x-axis at a point closer to the root than x.

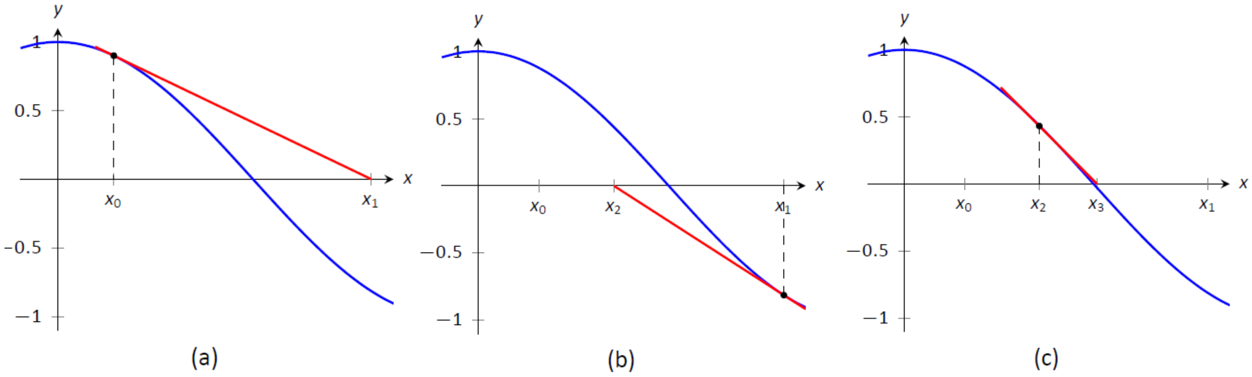


Figure $4.1.1$***:*** *Demonstrating the geometric concept behind Newton's Method*

We start Newton's Method with an initial guess about roughly where the root is. Call this x_0. (See Figure $4.1.1a$.) Draw the tangent line to the graph at $(x_0, f(x_0))$ and see where it meets the x-axis. Call this point x_1. Then repeat the process -- draw the tangent line to the graph at $(x_1, f(x_1))$ and see where it meets the x-axis. (See Figure $4.1.1b$.) Call this point x_2. Repeat the process again to get x_3, x_4, etc. This sequence of points will often converge rather quickly to a root of f.

We can use this *geometric* process to create an *algebraic* process. Let's look at how we found x_1. We started with the tangent line to the graph at $(x_0, f(x_0))$. The slope of this tangent line is $f'(x_0)$ and the equation of the line is

$$y=f'(x_0)(x-x_0)+f(x_0).\]

This line crosses the x-axis when $y=0$, and the x--value where it crosses is what we called x_1. So let $y=0$ and replace x with x_1, giving the equation:

$$0=f'(x_0)(x_1-x_0)+f(x_0). \tag{4.1.1}$$

Now solve for x_1:

$$x_1=x_0-\frac{f(x_0)}{f'(x_0)}. \tag{4.1.2}$$

Since we repeat the same geometric proces to find x_2 from x_1, we have

$$x_2=x_1-\frac{f(x_1)}{f'(x_1)}. \tag{4.1.3}$$

In general, given an approximation x_n, we can find the next approximation, x_{n+1} as follows:

$$x_{n+1}=x_n-\frac{f(x_n)}{f'(x_n)}. \tag{4.1.4}$$

We summarize this process as follows.

Key Idea 5: Newton's Method

Let f be a differentiable function on an interval I with a root in I. To approximate the value of the root, accurate to d decimal places:

1. Choose a value x_0 as an initial approximation of the root. (This is often done by looking at a graph of f.)
2. Create successive approximations iteratively; given an approximation x_n, compute the next approximation x_{n+1} as

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \tag{4.1.5}$$

3. Stop the iterations when successive approximations do not differ in the first d places after the decimal point.

Note: Newton's Method is not infallible. The sequence of approximate values may not converge, or it may converge so slowly that one is "tricked" into thinking a certain approximation is better than it actually is. These issues will be discussed at the end of the section.}

Let's practice Newton's Method with a concrete example.

Example 4.1.1: Using Newton's Method

Approximate the real root of $x^3-x^2-1=0$, accurate to the first 3 places after the decimal, using Newton's Method and an initial approximation of $x_0=1$.}

Solution

To begin, we compute $f'(x)=3x^2-2x$. Then we apply the Newton's Method algorithm, outlined in Key Idea 5.

$$x_1 = 1-\frac{f(1)}{f'(1)} = 1-\frac{1^3-1^2-1}{3\cdot 1^2-2\cdot 1} = 2, \tag{4.1.6}$$

$$x_2 = 2-\frac{f(2)}{f'(2)} = 2-\frac{2^3-2^2-1}{3\cdot 2^2-2\cdot 2} = 1.625, \tag{4.1.7}$$

$$x_3 = 1.625-\frac{f(1.625)}{f'(1.625)} = 1.625-\frac{1.625^3-1.625^2-1}{3\cdot 1.625^2-2\cdot 1.625} \approx 1.48579. \tag{4.1.8}$$

$$x_4 = 1.48579-\frac{f(1.48579)}{f'(1.48579)} \approx 1.46596 \tag{4.1.9}$$

$$x_5 = 1.46596-\frac{f(1.46596)}{f'(1.46596)} \approx 1.46557 \tag{4.1.10}$$

We performed 5 iterations of Newton's Method to find a root accurate to the first 3 places after the decimal; our final approximation is 1.465. The exact value of the root, to six decimal places, is 1.465571; It turns out that our x_5 is accurate to more than just 3 decimal places.

A graph of $f(x)$ is given in Figure $4.1.2$. We can see from the graph that our initial approximation of $x_0=1$ was not particularly accurate; a closer guess would have been $x_0=1.5$. Our choice was based on ease of initial calculation, and shows that Newton's Method can be robust enough that we do not have to make a very accurate initial approximation.

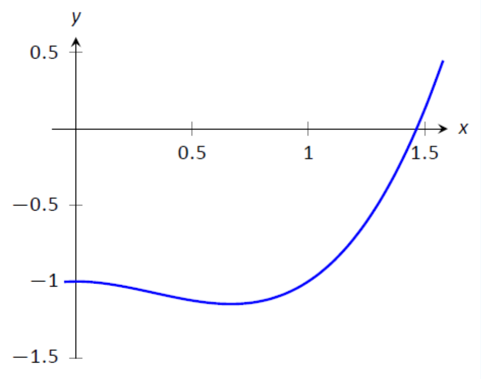


***Figure** $4.1.2$**:** A graph of $f(x)=x^3-x^2-1$ in Example $4.1.1$.*

We can automate this process on a calculator that has an *Ans* key that returns the result of the previous calculation. Start by pressing `1` and then `Enter`. (We have just entered our initial guess, $x_0 = 1$.) Now compute

$$\texttt{Ans} - \frac{f(\texttt{Ans})}{f'(\texttt{Ans})}\]

by entering the following and repeatedly press the `Enter` key:

$$\texttt{Ans} - (\texttt{Ans}^3 - \texttt{Ans}^2 - 1)/(3 * \texttt{Ans}^2 - 2 * \texttt{Ans}) \tag{4.1.11}$$

Each time we press the \texttt{Enter} key, we are finding the successive approximations, x_1, x_2, \dots, and each one is getting closer to the root. In fact, once we get past around x_7 or so, the approximations don't appear to be changing. They actually are changing, but the change is far enough to the right of the decimal point that it doesn't show up on the calculator's display. When this happens, we can be pretty confident that we have found an accurate approximation.

Using a calculator in this manner makes the calculations simple; many iterations can be computed very quickly.

Example 4.1.2: Using Newton's Method to find where functions intersect

Use Newton's Method to approximate a solution to $\cos x = x$, accurate to 5 places after the decimal.

Solution

Newton's Method provides a method of solving $f(x) = 0$; it is not (directly) a method for solving equations like $f(x) = g(x)$. However, this is not a problem; we can rewrite the latter equation as $f(x) - g(x) = 0$ and then use Newton's Method.

So we rewrite $\cos x = x$ as $\cos x - x = 0$. Written this way, we are finding a root of $f(x) = \cos x - x$. We compute $f'(x) = -\sin x - 1$. Next we need a starting value, x_0. Consider Figure 4.1.3, where $f(x) = \cos x - x$ is graphed. It seems that $x_0 = 0.75$ is pretty close to the root, so we will use that as our x_0. (The figure also shows the graphs of $y = \cos x$ and $y = x$, drawn with dashed lines. Note how they intersect at the same x value as when $f(x) = 0$.)

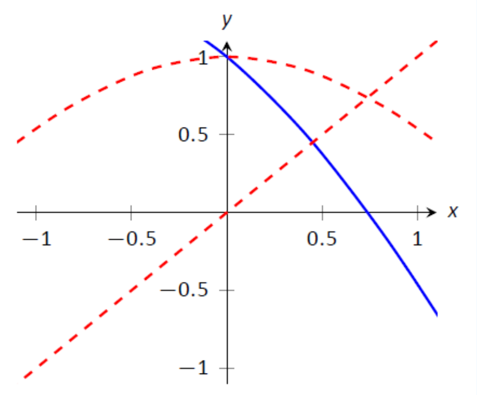


***Figure** 4.1.3: A graph of $f(x) = \cos x - x$ used to find an initial approximation of its root.*

We now compute x_1, x_2, etc. The formula for x_1 is

$$x_1 = 0.75 - \frac{\cos(0.75) - 0.75}{-\sin(0.75) - 1} \approx 0.7391111388. \tag{4.1.12}$$

Apply Newton's Method again to find x_2:

$$x_2 = 0.7391111388 - \frac{\cos(0.7391111388) - 0.7391111388}{-\sin(0.7391111388) - 1} \approx 0.7390851334. \tag{4.1.13}$$

We can continue this way, but it is really best to automate this process. On a calculator with an Ans key, we would start by pressing 0.75, then `Enter`, inputting our initial approximation. We then enter:

$$\texttt{Ans - (cos(Ans)-Ans)/(-sin(Ans)-1).} \tag{4.1.14}$$

Repeatedly pressing the `Enter` key gives successive approximations. We quickly find:

$$x_3 = 0.7390851332 \tag{4.1.15}$$
$$x_4 = 0.7390851332. \tag{4.1.16}$$

> Our approximations x_2 and x_3 did not differ for at least the first 5 places after the decimal, so we could have stopped. However, using our calculator in the manner described is easy, so finding x_4 was not hard. It is interesting to see how we found an approximation, accurate to as many decimal places as our calculator displays, in just 4 iterations.

If you know how to program, you can translate the following pseudocode into your favorite language to perform the computation in this problem.

```
x = .75
while true
    oldx = x
    x = x - (cos(x)-x)/(-sin(x)-1)
    print x
    if abs(x-oldx) < .0000000001
        break
```

This code calculates x_1, x_2, etc., storing each result in the variable `x`. The previous approximation is stored in the variable `oldx`. We continue looping until the difference between two successive approximations, `abs(x-oldx)`, is less than some small tolerance, in this case, `.0000000001`.

Convergence of Newton's Method

What should one use for the initial guess, x_0? Generally, the closer to the actual root the initial guess is, the better. However, some initial guesses should be avoided. For instance, consider Example $4.1.1$ where we sought the root to $f(x) = x^3 - x^2 - 1$. Choosing $x_0 = 0$ would have been a particularly poor choice. Consider Figure $4.1.4$, where $f(x)$ is graphed along with its tangent line at $x = 0$. Since $f'(0) = 0$, the tangent line is horizontal and does not intersect the x--axis. Graphically, we see that Newton's Method fails.

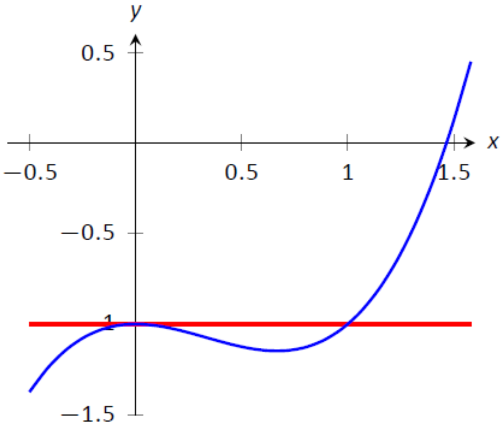


***Figure** $4.1.4$**:** A graph of $f(x) = x^3 - x^2 - 1$, showing why an initial approximation of $x_0 = 0$ with Newton's Method fails.*

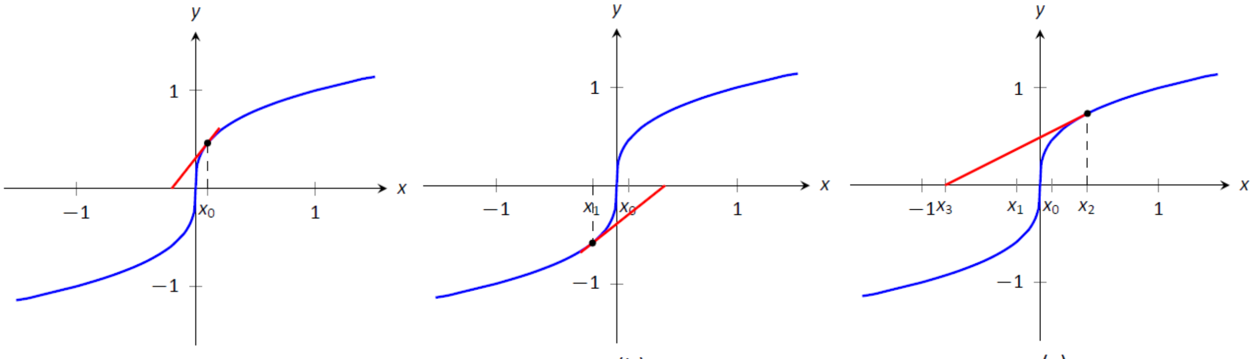
We can also see analytically that it fails. Since

$$x_1 = 0 - \frac{f(0)}{f'(0)} \tag{4.1.17}$$

and $f'(0) = 0$, we see that x_1 is not well defined.

This problem can also occur if, for instance, it turns out that $f'(x_5) = 0$. Adjusting the initial approximation x_0 by a very small amount will likely fix the problem.

It is also possible for Newton's Method to not converge while each successive approximation is well defined. Consider $f(x) = x^{1/3}$, as shown in Figure $4.1.5$. It is clear that the root is $x = 0$, but let's approximate this with $x_0 = 0.1$. Figure $4.1.5a$ shows graphically the calculation of x_1; notice how it is farther from the root than x_0. Figures $4.1.5b$ and $(4.1.5c)$ show the calculation of x_2 and x_3, which are even farther away; our successive approximations are getting worse. (It turns out that in this particular example, each successive approximation is twice as far from the true answer as the previous approximation.)



(a) (b) (c)

***Figure** 4.1.5**:** Newton's Method fails to find a root of $f(x) = x^{1/3}$, regardless of the choice of x_0.}\label{fig:newt4}*

There is no "fix" to this problem; Newton's Method simply will not work and another method must be used.

While Newton's Method does not always work, it does work "most of the time," and it is generally very fast. Once the approximations get close to the root, Newton's Method can as much as double the number of correct decimal places with each successive approximation. A course in Numerical Analysis will introduce the reader to more iterative root finding methods, as well as give greater detail about the strengths and weaknesses of Newton's Method.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/
- Integrated by Justin Marshall.

This page titled 4.1: Newton's Method is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

4.2: Related Rates

When two quantities are related by an equation, knowing the value of one quantity can determine the value of the other. For instance, the circumference and radius of a circle are related by $C = 2\pi r$; knowing that $C = 6\pi$ in determines the radius must be 3 in.

The topic of **related rates** takes this one step further: knowing the *rate* at which one quantity is changing can determine the *rate* at which the other changes.

Note: This section relies heavily on implicit differentiation, so referring back to Section 2.6 may help. We demonstrate the concepts of related rates through examples.

Example 4.2.1: Understanding related rates

The radius of a circle is growing at a rate of 5 in/hr. At what rate is the circumference growing?

Solution

The circumference and radius of a circle are related by $C = 2\pi r$. We are given information about how the length of r changes with respect to time; that is, we are told $\frac{dr}{dt} = 5$in/hr. We want to know how the length of C changes with respect to time, i.e., we want to know $\frac{dC}{dt}$.

Implicitly differentiate both sides of $C = 2\pi r$ with respect to t:

$$\begin{aligned} C &= 2\pi r \\ \frac{d}{dt}\big(C\big) &= \frac{d}{dt}\big(2\pi r\big) \\ \frac{dC}{dt} &= 2\pi\frac{dr}{dt}. \end{aligned}$$

As we know $\frac{dr}{dt} = 5$in/hr, we know

$$\frac{dC}{dt} = 2\pi 5 = 10\pi \approx 31.4\text{in/hr.}\]

Consider another, similar example.

Example 4.2.2: Finding related rates

Water streams out of a faucet at a rate of $2\text{in}^3/\text{s}$ onto a flat surface at a constant rate, forming a circular puddle that is $1/8$in deep.

1. At what rate is the area of the puddle growing?
2. At what rate is the radius of the circle growing?

Solution

1. We can answer this question two ways: using "common sense" or related rates. The common sense method states that the volume of the puddle is growing by $2\text{in}^3/\text{s}$, where

$$\text{volume of puddle} \;=\; \text{area of circle} \times \text{depth}. \tag{4.2.1}$$

Since the depth is constant at $1/8$in, the area must be growing by $16\text{in}^2/\text{s}$.

This approach reveals the underlying related--rates principle. Let V and A represent the Volume and Area of the puddle. We know $V = A \times \frac{1}{8}$. Take the derivative of both sides with respect to t, employing implicit differentiation.

$$V = \frac{1}{8}A \tag{4.2.2}$$

$$\frac{d}{dt}\big(V\big) = \frac{d}{dt}\left(\frac{1}{8}A\right) \tag{4.2.3}$$

$$\frac{dV}{dt} = \frac{1}{8}\frac{dA}{dt} \tag{4.2.4}$$

As $\frac{dV}{dt} = 2$, we know $2 = \frac{1}{8}\frac{dA}{dt}$, and hence $\frac{dA}{dt} = 16$. Thus the area is growing by $16\text{in}^2/\text{s}$.

2. To start, we need an equation that relates what we know to the radius. We just learned something about the surface area of the circular puddle, and we know $A = \pi r^2$. We should be able to learn about the rate at which the radius is growing with this information.

Implicitly derive both sides of $A = \pi r^2$ with respect to t:

$$A = \pi r^2 \tag{4.2.5}$$

$$\frac{d}{dt}\left(A\right) = \frac{d}{dt}\left(\pi r^2\right) \tag{4.2.6}$$

$$\frac{dA}{dt} = 2\pi r\frac{dr}{dt} \tag{4.2.7}$$

Our work above told us that $\frac{dA}{dt} = 16\text{in}^2$/s. Solving for $\frac{dr}{dt}$, we have

$$\frac{dr}{dt} = \frac{8}{\pi r}.$$

Note how our answer is not a number, but rather a function of r. In other words, *the rate at which the radius is growing depends on how big the circle already is.* If the circle is very large, adding 2in^3 of water will not make the circle much bigger at all. If the circle dime--sized, adding the same amount of water will make a radical change in the radius of the circle.

In some ways, our problem was (intentionally) ill--posed. We need to specify a current radius in order to know a rate of change. When the puddle has a radius of 10in, the radius is growing at a rate of

$$\frac{dr}{dt} = \frac{8}{10\pi} = \frac{4}{5\pi} \approx 0.25\text{in/s}.$$

Example $4.2.3$: Studying related rates

Radar guns measure the rate of distance change between the gun and the object it is measuring. For instance, a reading of "55mph" means the object is moving away from the gun at a rate of 55 miles per hour, whereas a measurement of "-25mph" would mean that the object is approaching the gun at a rate of 25 miles per hour.

If the radar gun is moving (say, attached to a police car) then radar readouts are only immediately understandable if the gun and the object are moving along the same line. If a police officer is traveling 60mph and gets a readout of 15mph, he knows that the car ahead of him is moving away at a rate of 15 miles an hour, meaning the car is traveling 75mph. (This straight--line principle is one reason officers park on the side of the highway and try to shoot straight back down the road. It gives the most accurate reading.)

Suppose an officer is driving due north at 30 mph and sees a car moving due east, as shown in Figure $4.2.1$. Using his radar gun, he measures a reading of 20mph. By using landmarks, he believes both he and the other car are about 1/2 mile from the intersection of their two roads.

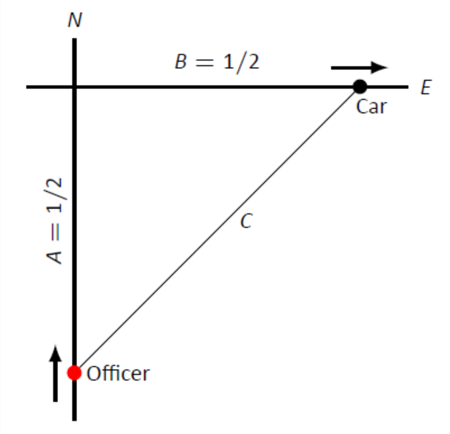


***Figure** $4.2.1$**:** A sketch of a police car (at bottom) attempting to measure the speed of a car (at right) in Example $4.2.3$.*

If the speed limit on the other road is 55mph, is the other driver speeding?

Solution

Using the diagram in Figure $4.2.1$, let's label what we know about the situation. As both the police officer and other driver are 1/2 mile from the intersection, we have $A = 1/2$, $B = 1/2$, and through the Pythagorean Theorem, $C = 1/\sqrt{2} \approx 0.707$.

We know the police officer is traveling at 30mph; that is, $\frac{dA}{dt} = -30$. The reason this rate of change is negative is that A is getting smaller; the distance between the officer and the intersection is shrinking. The radar measurement is $\frac{dC}{dt} = 20$. We want to find $\frac{dB}{dt}$.

We need an equation that relates B to A and/or C. The Pythagorean Theorem is a good choice: $A^2 + B^2 = C^2$. Differentiate both sides with respect to t:

$$A^2 + B^2 = C^2 \tag{4.2.8}$$

$$\frac{d}{dt}\left(A^2 + B^2\right) = \frac{d}{dt}\left(C^2\right) \tag{4.2.9}$$

$$2A\frac{dA}{dt} + 2B\frac{dB}{dt} = 2C\frac{dC}{dt} \tag{4.2.10}$$

We have values for everything except $\frac{dB}{dt}$. Solving for this we have

$$\frac{dB}{dt} = \frac{C\frac{dC}{dt}- A\frac{dA}{dt}}{B} \approx 58.28\text{mph}.\]

The other driver appears to be speeding slightly.

Note: Example $4.2.3$ is both interesting and impractical. It highlights the difficulty in using radar in a non--linear fashion, and explains why "in real life" the police officer would follow the other driver to determine their speed, and not pull out pencil and paper.

The principles here are important, though. Many automated vehicles make judgments about other moving objects based on perceived distances, radar--like measurements and the concepts of related rates.

Example $4.2.4$: Studying related rates

A camera is placed on a tripod 10ft from the side of a road. The camera is to turn to track a car that is to drive by at 100mph for a promotional video. The video's planners want to know what kind of motor the tripod should be equipped with in order to properly track the car as it passes by. Figure $4.2.2$ shows the proposed setup.

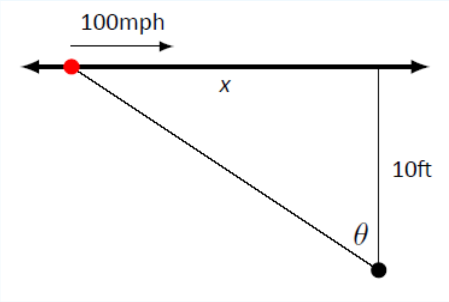


***Figure** $4.2.2$**:** Tracking a speeding car (at left) with a rotating camera.*

How fast must the camera be able to turn to track the car?

Solution

We seek information about how fast the camera is to *turn*; therefore, we need an equation that will relate an angle θ to the position of the camera and the speed and position of the car.

Figure $4.2.2$ suggests we use a trigonometric equation. Letting x represent the distance the car is from the point on the road directly in front of the camera, we have

$$\tan\theta = \frac{x}{10} \tag{4.2.11}$$

As the car is moving at 100mph, we have $\frac{dx}{dt} = -100$mph (as in the last example, since x is getting smaller as the car travels, $\frac{dx}{dt}$ is negative). We need to convert the measurements so they use the same units; rewrite -100mph in terms of ft/s:

$$\frac{dx}{dt} = -100\frac{\text{m}}{\text{hr}} = -100\frac{\text{m}}{\text{hr}}\cdot5280\frac{\text{ft}}{\text{m}}\cdot\frac{1}{3600}\frac{\text{hr}}{\text{s}} =-146.\overline{6}\text{ft/s}.\]

Now take the derivative of both sides of Equation $4.2.9$ using implicit differentiation:

$$\tan\theta = \frac{x}{10} \tag{4.2.12}$$

$$\frac{d}{dt}\big(\tan\theta\big) = \frac{d}{dt}\left(\frac{x}{10}\right) \tag{4.2.13}$$

$$\sec^2\theta\,\frac{d\theta}{dt} = \frac{1}{10}\,\frac{dx}{dt} \tag{4.2.14}$$

$$\frac{d\theta}{dt} = \frac{\cos^2\theta}{10}\,\frac{dx}{dt} \tag{4.2.15}$$

We want to know the fastest the camera has to turn. Common sense tells us this is when the car is directly in front of the camera (i.e., when $\theta = 0$). Our mathematics bears this out. In Equation $4.2.14$ we see this is when $\cos^2\theta$ is largest; this is when $\cos\theta = 1$, or when $\theta = 0$.

With $\frac{dx}{dt} \approx -146.67$ft/s, we have

$$\frac{d\theta}{dt} = -\frac{1\text{rad}}{10\text{ft}}146.67\text{ft/s} = -14.667\text{radians/s}.\]

We find that $\frac{d\theta}{dt}$ is negative; this matches our diagram in Figure $4.2.2$ for θ is getting smaller as the car approaches the camera.

What is the practical meaning of -14.667 radians/s? Recall that 1 circular revolution goes through 2π radians, thus 14.667 rad/s means $14.667/(2\pi) \approx 2.33$ revolutions per second. The negative sign indicates the camera is rotating in a clockwise fashion.

We introduced the derivative as a function that gives the slopes of tangent lines of functions. This chapter emphasizes using the derivative in other ways. Newton's Method uses the derivative to approximate roots of functions; this section stresses the "rate of change" aspect of the derivative to find a relationship between the rates of change of two related quantities. In the next section we use Extreme Value concepts to *optimize* quantities.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/
- Integrated by Justin Marshall.

This page titled 4.2: Related Rates is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

4.3: Optimization

In Section 3.1 we learned about extreme values -- the largest and smallest values a function attains on an interval. We motivated our interest in such values by discussing how it made sense to want to know the highest/lowest values of a stock, or the fastest/slowest an object was moving. In this section we apply the concepts of extreme values to solve "word problems," i.e., problems stated in terms of situations that require us to create the appropriate mathematical framework in which to solve the problem.

We start with a classic example which is followed by a discussion of the topic of optimization.

Example $4.3.1$: Optimization: perimeter and area

A man has 100 feet of fencing, a large yard, and a small dog. He wants to create a rectangular enclosure for his dog with the fencing that provides the maximal area. What dimensions provide the maximal area?

Solution

One can likely guess the correct answer -- that is great. We will proceed to show how calculus can provide this answer in a context that proves this answer is correct.

It helps to make a sketch of the situation. Our enclosure is sketched twice in Figure $4.3.1$, either with green grass and nice fence boards or as a simple rectangle. Either way, drawing a rectangle forces us to realize that we need to know the dimensions of this rectangle so we can create an area function -- after all, we are trying to maximize the area.

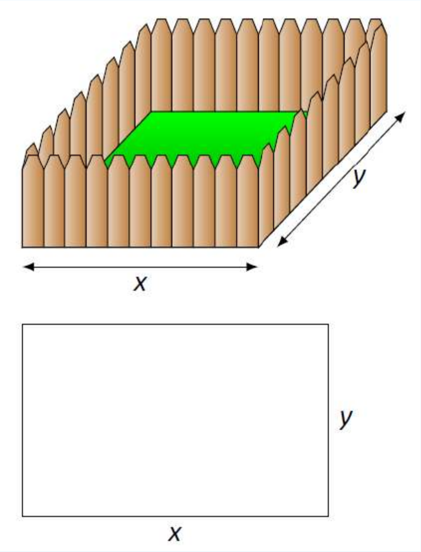


***Figure** $4.3.1$**:** A sketch of the enclosure in Example* $4.3.1$.

We let x and y denote the lengths of the sides of the rectangle. Clearly,

$$\text{Area} = xy. \tag{4.3.1}$$

We do not yet know how to handle functions with 2 variables; we need to reduce this down to a single variable. We know more about the situation: the man has 100 feet of fencing. By knowing the perimeter of the rectangle must be 100, we can create another equation:

$$\text{Perimeter} = 100 = 2x + 2y. \tag{4.3.2}$$

We now have 2 equations and 2 unknowns. In the latter equation, we solve for y:

$$y = 50 - x. \tag{4.3.3}$$

Now substitute this expression for y in the area equation:

$$\text{Area} = A(x) = x(50 - x). \tag{4.3.4}$$

Note we now have an equation of one variable; we can truly call the Area a function of x.

This function only makes sense when $0 \leq x \leq 50$, otherwise we get negative values of area. So we find the extreme values of $A(x)$ on the interval $[0, 50]$.

To find the critical points, we take the derivative of $A(x)$ and set it equal to 0, then solve for x.

$$\begin{align} A(x) &= x(50-x) \tag{4.3.5} \\ &= 50x - x^2 \tag{4.3.6} \\ A'(x) &= 50 - 2x \tag{4.3.7} \end{align}$$

We solve $50 - 2x = 0$ to find $x = 25$; this is the only critical point. We evaluate $A(x)$ at the endpoints of our interval and at this critical point to find the extreme values; in this case, all we care about is the maximum.

Clearly $A(0) = 0$ and $A(50) = 0$, whereas $A(25) = 625\text{ft}^2$. This is the maximum. Since we earlier found $y = 50 - x$, we find that y is also 25. Thus the dimensions of the rectangular enclosure with perimeter of 100 ft. with maximum area is a square, with sides of length 25 ft.

This example is very simplistic and a bit contrived. (After all, most people create a design then buy fencing to meet their needs, and not buy fencing and plan later.) But it models well the necessary process: create equations that describe a situation, reduce an equation to a single variable, then find the needed extreme value.

"In real life," problems are much more complex. The equations are often *not* reducible to a single variable (hence multi--variable calculus is needed) and the equations themselves may be difficult to form. Understanding the principles here will provide a good foundation for the mathematics you will likely encounter later.

We outline here the basic process of solving these optimization problems.

Key Idea 6: Solving Optimization Problems

1. Understand the problem. Clearly identify what quantity is to be maximized or minimized. Make a sketch if helpful.
2. Create equations relevant to the context of the problem, using the information given. (One of these should describe the quantity to be optimized. We'll call this the *fundamental equation*.)
3. If the fundamental equation defines the quantity to be optimized as a function of more than one variable, reduce it to a single variable function using substitutions derived from the other equations.
4. Identify the domain of this function, keeping in mind the context of the problem.
5. Find the extreme values of this function on the determined domain.
6. Identify the values of all relevant quantities of the problem.

We will use Key Idea 6 in a variety of examples.

Example $4.3.2$: {Optimization: perimeter and area

Here is another classic calculus problem: A woman has a 100 feet of fencing, a small dog, and a large yard that contains a stream (that is mostly straight). She wants to create a rectangular enclosure with maximal area that uses the stream as one side. (Apparently her dog won't swim away.) What dimensions provide the maximal area?

Solution

We will follow the steps outlined by Key Idea 6.

1. We are maximizing *area*. A sketch of the region will help; Figure $4.3.2$ gives two sketches of the proposed enclosed area. A key feature of the sketches is to acknowledge that one side is not fenced.

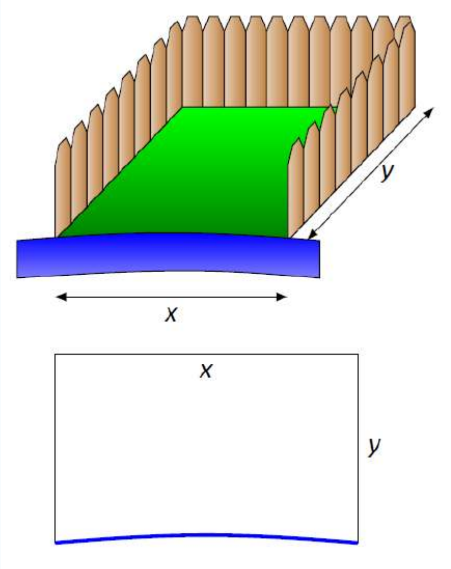


Figure 4.3.2: A sketch of the enclosure in Example 4.3.2.

2. We want to maximize the area; as in the example before,

$$\text{Area} = xy. \tag{4.3.8}$$

   This is our fundamental equation. This defines area as a function of two variables, so we need another equation to reduce it to one variable.
   We again appeal to the perimeter; here the perimeter is

$$\text{Perimeter} = 100 = x + 2y. \tag{4.3.9}$$

   Note how this is different than in our previous example.
3. We now reduce the fundamental equation to a single variable. In the perimeter equation, solve for y: $y = 50 - x/2$. We can now write Area as

$$\text{Area} = A(x) = x(50 - x/2) = 50x - \frac{1}{2}x^2. \tag{4.3.10}$$

   Area is now defined as a function of one variable.
4. We want the area to be nonnegative. Since $A(x) = x(50 - x/2)$, we want $x \geq 0$ and $50 - x/2 \geq 0$. The latter inequality implies that $x \leq 100$, so $0 \leq x \leq 100$.
5. We now find the extreme values. At the endpoints, the minimum is found, giving an area of 0.
   Find the critical points. We have $A'(x) = 50 - x$; setting this equal to 0 and solving for x returns $x = 50$. This gives an area of

$$A(50) = 50(25) = 1250. \tag{4.3.11}$$

6. We earlier set $y = 50 - x/2$; thus $y = 25$. Thus our rectangle will have two sides of length 25 and one side of length 50, with a total area of 1250 ft^2.

Keep in mind as we do these problems that we are practicing a *process*; that is, we are learning to turn a situation into a system of equations. These equations allow us to write a certain quantity as a function of one variable, which we then optimize.

Example $4.3.3$: Optimization: minimizing cost

A power line needs to be run from an power station located on the beach to an offshore facility. Figure $4.3.3$ shows the distances between the power station to the facility.

It costs \$50/ft. to run a power line along the land, and \$130/ft. to run a power line under water. How much of the power line should be run along the land to minimize the overall cost? What is the minimal cost?

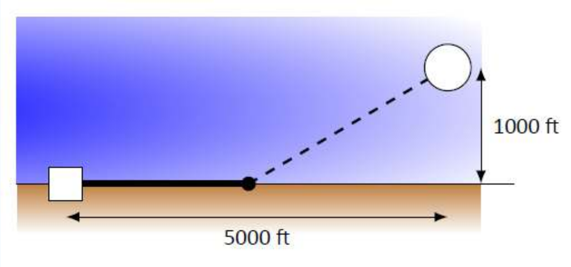


Figure $4.3.3$***:*** *Running a power line from the power station to an offshore facility with minimal cost in Example* $4.3.3$

Solution

We will follow the strategy of Key Idea 6 implicitly, without specifically numbering steps.

There are two immediate solutions that we could consider, each of which we will reject through "common sense." First, we could minimize the distance by directly connecting the two locations with a straight line. However, this requires that all the wire be laid underwater, the most costly option. Second, we could minimize the underwater length by running a wire all 5000 ft. along the beach, directly across from the offshore facility. This has the undesired effect of having the longest distance of all, probably ensuring a non--minimal cost.

The optimal solution likely has the line being run along the ground for a while, then underwater, as the figure implies. We need to label our unknown distances -- the distance run along the ground and the distance run underwater. Recognizing that the underwater distance can be measured as the hypotenuse of a right triangle, we choose to label the distances as shown in Figure $4.3.4$.

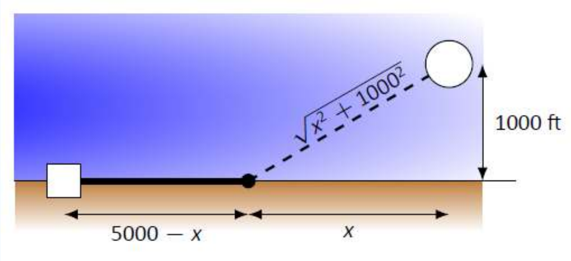


Figure $4.3.4$***:*** *Labeling unknown distances in Example* $4.3.3$.

By choosing x as we did, we make the expression under the square root simple. We now create the cost function.

$$\begin{aligned} \text{Cost} &= \text{land cost} &&+ \text{water cost} \\ &\$50\times\text{land distance} &&+ \$130\times\text{water distance} \\ &50(5000-x) &&+ 130\sqrt{x^2+1000^2}. \end{aligned} \tag{4.3.12}$$

So we have $c(x)=50(5000-x)+130\sqrt{x^2+1000^2}$. This function only makes sense on the interval $[0,5000]$. While we are fairly certain the endpoints will not give a minimal cost, we still evaluate $c(x)$ at each to verify.

$$c(0) = 380,000 \quad\quad c(5000) \approx 662,873.$$

We now find the critical values of $c(x)$. We compute $c'(x)$ as

$$c'(x) = -50+\frac{130x}{\sqrt{x^2+1000^2}}.$$

Recognize that this is never undefined. Setting $c'(x)=0$ and solving for x, we have:

$$-50+\frac{130x}{\sqrt{x^2+1000^2}}=0 \tag{4.3.13}$$

$$\frac{130x}{\sqrt{x^2+1000^2}}=50 \tag{4.3.14}$$

$$\frac{130^2x^2}{x^2+1000^2} = 50^2 \tag{4.3.15}$$

$$130^2x^2 = 50^2(x^2+1000^2) \tag{4.3.16}$$

$$130^2x^2 - 50^2x^2 = 50^2 \cdot 1000^2 \tag{4.3.17}$$

$$(130^2-50^2)x^2 = 50,000^2 \tag{4.3.18}$$

$$x^2 = \frac{50,000^2}{130^2-50^2} \tag{4.3.19}$$

$$x = \frac{50,000}{\sqrt{130^2-50^2}} \tag{4.3.20}$$

$$x = \frac{50,000}{120} = 416\frac{2}{3} \approx 416.67. \tag{4.3.21}$$

Evaluating $c(x)$ at $x = 416.67$ gives a cost of about $370,000. The distance the power line is laid along land is $5000 - 416.67 = 4583.33$ ft., and the underwater distance is $\sqrt{416.67^2+1000^2} \approx 1083$ ft.

In the exercises you will see a variety of situations that require you to combine problem--solving skills with calculus. Focus on the *process*; learn how to form equations from situations that can be manipulated into what you need. Eschew memorizing how to do "this kind of problem" as opposed to "that kind of problem." Learning a process will benefit one far longer than memorizing a specific technique.

The next section introduces our final application of the derivative: *differentials*. Given $y = f(x)$, they offer a method of approximating the change in y after x changes by a small amount.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/
- Integrated by Justin Marshall.

This page titled 4.3: Optimization is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

4.4: Differentials

In Section 2.2 we explored the meaning and use of the derivative. This section starts by revisiting some of those ideas.

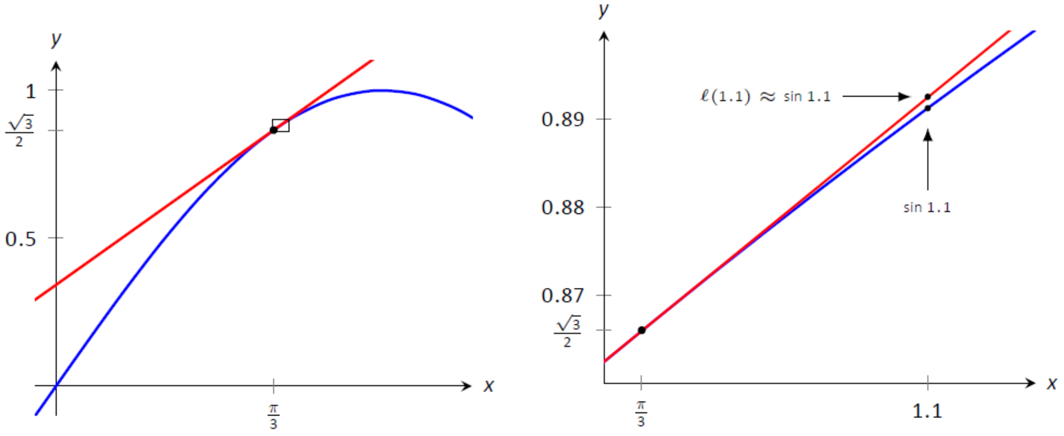
Recall that the derivative of a function f can be used to find the slopes of lines tangent to the graph of f. At $x=c$, the tangent line to the graph of f has equation

$$y = f'(c)(x-c)+f(c).$$

The tangent line can be used to find good approximations of $f(x)$ for values of x near c.

For instance, we can approximate $\sin 1.1$ using the tangent line to the graph of $f(x)=\sin x$ at $x=\pi/3\approx 1.05$. Recall that $\sin(\pi/3) = \sqrt{3}/2\approx 0.866$, and $\cos(\pi/3)=1/2$. Thus the tangent line to $f(x)=\sin x$ at $x=\pi/3$ is:

$$\ell(x) = \frac12(x-\pi/3)+0.866.$$



(a) (b)

***Figure** 4.4.1: Graphing $f(x)=\sin x$ and its tangent line at $x=\pi/3$ in order to estimate $\sin 1.1$.*

In Figure $4.4.1a$, we see a graph of $f(x)=\sin x$ graphed along with its tangent line at $x=\pi/3$. The small rectangle shows the region that is displayed in Figure $4.4.1b$. In this figure, we see how we are approximating $\sin 1.1$ with the tangent line, evaluated at 1.1. Together, the two figures show how close these values are.

Using this line to approximate $\sin 1.1$, we have:

$$\begin{aligned}\ell(1.1) &= \frac{1}{2}(1.1-\pi/3)+0.866 && (4.4.1)\\ &= \frac{1}{2}(0.053)+0.866 = 0.8925. && (4.4.2)\end{aligned}$$

(We leave it to the reader to see how good of an approximation this is.)

We now generalize this concept. Given $f(x)$ and an x--value c, the tangent line is $\ell(x)=f'(c)(x-c)+f(c)$. Clearly, $f(c)=\ell(c)$. Let Δx be a small number, representing a small change in x value. We assert that:

$$f(c+\Delta x) \approx \ell(c+\Delta x),$$

since the tangent line to a function approximates well the values of that function near $x=c$.

As the x value changes from c to $c+\Delta x$, the y value of f changes from $f(c)$ to $f(c+\Delta x)$. We call this change of y value Δy. That is:

$$\Delta y = f(c+\Delta x)-f(c).$$

Replacing $f(c+\Delta x)$ with its tangent line approximation, we have

$$\begin{aligned}\Delta y &\approx \ell(c+\Delta x)-f(c)\\ &= f'(c)\big((c+\Delta x)-c\big)+f(c)-f(c)\\ &= f'(c)\Delta x && (4.4.3)\end{aligned}$$

This final equation is important; we'll come back to it in Key Idea 7.

We introduce two new variables, dx and dy in the context of a formal definition.

Definition: Differentials of x and y.

Let $y = f(x)$ be differentiable. The *differential* of x, denoted dx, is any nonzero real number (usually taken to be a small number). The *differential* of y, denoted dy, is

$$dy = f'(x)dx. \tag{4.4.4}$$

We can solve for $f'(x)$ in the above equation: $f'(x) = dy/dx$. This states that the derivative of f with respect to x is the differential of y divided by the differential of x; this is **not** the alternate notation for the derivative, $\frac{dy}{dx}$. This latter notation was chosen because of the fraction--like qualities of the derivative, but again, it is one symbol and not a fraction.

It is helpful to organize our new concepts and notations in one place.

Key Idea 7: Differential Notation

Let $y = f(x)$ be a differentiable function.

1. Δx represents a small, nonzero change in x value.
2. dx represents a small, nonzero change in x value (i.e., $\Delta x = dx$).
3. Δy is the change in y value as x changes by Δx; hence

$$\Delta y = f(x + \Delta x) - f(x). \tag{4.4.5}$$

4. $dy = f'(x)dx$ which, by Equation 4.4.7, is an *approximation* of the change in y value as x changes by Δx; $dy \approx \Delta y$.

What is the value of differentials? Like many mathematical concepts, differentials provide both practical and theoretical benefits. We explore both here.

Example 4.4.1: Finding and using differentials

Consider $f(x) = x^2$. Knowing $f(3) = 9$, approximate $f(3.1)$.

Solution

The x value is changing from $x = 3$ to $x = 3.1$; therefore, we see that $dx = 0.1$. If we know how much the y value changes from $f(3)$ to $f(3.1)$ (i.e., if we know Δy), we will know exactly what $f(3.1)$ is (since we already know $f(3)$). We can approximate \Delta y\ with dy.

$$\begin{align} \Delta y &\approx dy \tag{4.4.6} \\ &= f'(3)dx \tag{4.4.7} \\ &= 2 \cdot 3 \cdot 0.1 = 0.6. \tag{4.4.8} \end{align}$$

We expect the y value to change by about 0.6, so we approximate $f(3.1) \approx 9.6.$

We leave it to the reader to verify this, but the preceding discussion links the differential to the tangent line of $f(x)$ at $x = 3$. One can verify that the tangent line, evaluated at $x = 3.1$, also gives $y = 9.6$.

Of course, it is easy to compute the actual answer (by hand or with a calculator): $3.1^2 = 9.61.$ (Before we get too cynical and say "Then why bother?", note our approximation is *really* good!)

So why bother?

In "most" real life situations, we do not know the function that describes a particular behavior. Instead, we can only take measurements of how things change -- measurements of the derivative.

Imagine water flowing down a winding channel. It is easy to measure the speed and direction (i.e., the *velocity*) of water at any location. It is very hard to create a function that describes the overall flow, hence it is hard to predict where a floating object placed at the beginning of the channel will end up. However, we can *approximate* the path of an object using differentials. Over small intervals, the path taken by a floating object is essentially linear. Differentials allow us to approximate the true path by piecing together lots of short, linear paths. This technique is called Euler's Method, studied in introductory Differential Equations courses.

We use differentials once more to approximate the value of a function. Even though calculators are very accessible, it is neat to see how these techniques can sometimes be used to easily compute something that looks rather hard.

Example $4.4.2$: Using differentials to approximate a function value

Approximate $\sqrt{4.5}$.

Solution

We expect $\sqrt{4.5} \approx 2$, yet we can do better. Let $f(x) = \sqrt{x}$, and let $c = 4$. Thus $f(4) = 2$. We can compute $f'(x) = 1/(2\sqrt{x})$, so $f'(4) = 1/4$.

We approximate the difference between $f(4.5)$ and $f(4)$ using differentials, with $dx = 0.5$:

$$f(4.5)-f(4) = \Delta y \approx dy = f'(4)\cdot dx = 1/4 \cdot 1/2 = 1/8 = 0.125.\]

The approximate change in f from $x = 4$ to $x = 4.5$ is 0.125, so we approximate $\sqrt{4.5} \approx 2.125.$

Differentials are important when we discuss *integration*. When we study that topic, we will use notation such as

$$\int f(x)\ dx\]

quite often. While we don't discuss here what all of that notation means, note the existence of the differential dx. Proper handling of *integrals* comes with proper handling of differentials.

In light of that, we practice finding differentials in general.

Example $4.4.3$: Finding differentials

In each of the following, find the differential dy.

$$y = \sin x \qquad 2. y = e^x(x^2+2) \qquad 3. y = \sqrt{x^2+3x-1} \tag{4.4.9}$$

Solution

1. $y = \sin x$: As $f(x) = \sin x$, $f'(x) = \cos x$. Thus

$$dy = \cos(x)dx. \tag{4.4.10}$$

2. $y = e^x(x^2+2)$: Let $f(x) = e^x(x^2+2)$. We need $f'(x)$, requiring the Product Rule.
   We have $f'(x) = e^x(x^2+2) + 2xe^x$, so

$$dy = \left(e^x(x^2+2) + 2xe^x\right)dx. \tag{4.4.11}$$

3. $y = \sqrt{x^2+3x-1}$: Let $f(x) = \sqrt{x^2+3x-1}$; we need $f'(x)$, requiring the Chain Rule.
   We have $\Delta s f'(x) = \frac{1}{2}(x^2+3x-1)^{-\frac{1}{2}}(2x+3) = \frac{2x+3}{2\sqrt{x^2+3x-1}}$. Thus

$$dy = \frac{(2x+3)dx}{2\sqrt{x^2+3x-1}}. \tag{4.4.12}$$

Finding the differential dy of $y = f(x)$ is really no harder than finding the derivative of f; we just *multiply* $f'(x)$ by dx. It is important to remember that we are not simply adding the symbol "dx" at the end.

We have seen a practical use of differentials as they offer a good method of making certain approximations. Another use is *error propagation*. Suppose a length is measured to be x, although the actual value is $x + \Delta x$ (where we hope \Delta x\ is small). This measurement of x may be used to compute some other value; we can think of this as $f(x)$ for some function f. As the true length is $x + \Delta x$, one really should have computed $f(x + \Delta x)$. The difference between $f(x)$ and $f(x + \Delta x)$ is the propagated error.

How close are $f(x)$ and $f(x + \Delta x)$? This is a difference in "y" values;

$$f(x+\Delta x)-f(x) = \Delta y \approx dy.\]

We can approximate the propagated error using differentials.

Example $4.4.4$: Using differentials to approximate propagated error

A steel ball bearing is to be manufactured with a diameter of 2cm. The manufacturing process has a tolerance of ± 0.1mm in the diameter. Given that the density of steel is about 7.85g/cm^3, estimate the propagated error in the mass of the ball bearing.

Solution

The mass of a ball bearing is found using the equation "mass = volume $\times$ density." In this situation the mass function is a product of the radius of the ball bearing, hence it is $m = 7.85\frac{4}{3}\pi r^3$. The differential of the mass is

$$dm = 31.4\pi r^2 dr.\]

The radius is to be 1cm; the manufacturing tolerance in the radius is ± 0.05mm, or ± 0.005cm. The propagated error is approximately:

$$\begin{aligned} \Delta m &\approx dm && (4.4.13)\\ &= 31.4\pi(1)^2(\pm 0.005) && (4.4.14)\\ &= \pm 0.493\text{g} && (4.4.15) \end{aligned}$$

Is this error significant? It certainly depends on the application, but we can get an idea by computing the *relative error*. The ratio between amount of error to the total mass is

$$\begin{aligned} \frac{dm}{m} &= \pm\frac{0.493}{7.85\frac{4}{3}\pi} && (4.4.16)\\ &= \pm\frac{0.493}{32.88} && (4.4.17)\\ &= \pm 0.015, && (4.4.18) \end{aligned}$$

or ± 1.5.

We leave it to the reader to confirm this, but if the diameter of the ball was supposed to be 10cm, the same manufacturing tolerance would give a propagated error in mass of ± 12.33g, which corresponds to a \textit{percent error} of ± 0.188\%. While the amount of error is much greater ($12.33 > 0.493$), the percent error is much lower.

We first learned of the derivative in the context of instantaneous rates of change and slopes of tangent lines. We furthered our understanding of the power of the derivative by studying how it relates to the graph of a function (leading to ideas of increasing/decreasing and concavity). This chapter has put the derivative to yet more uses:

- Equation solving (Newton's Method)
- Related Rates (furthering our use of the derivative to find instantaneous rates of change)
- Optimization (applied extreme values), and
- Differentials (useful for various approximations and for something called integration).

In the next chapters, we will consider the "reverse" problem to computing the derivative: given a function f, can we find a function whose derivative is f? Be able to do so opens up an incredible world of mathematics and applications.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/
- Integrated by Justin Marshall.

This page titled 4.4: Differentials is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

4.E: Applications of Derivatives (Exercises)

4.1: Newton's Method

Terms and Concepts

1. T/F: Given a function $f(x)$, Newton's Method produces an exact solution to $f(x)=0$.

2. T/F: In order to get a solution to $f(x)=0$ accurate to d places after the decimal, at least $d+1$ iterations of Newton's Method must be used.

Problems

In Exercises 3-7, the roots of $f(x)$ are known or are easily found. Use 5 iterations of Newton's Method with the given initial approximation to approximate the root. Compare it to the known value of the root.

3. $f(x)=\cos x,\ x_0=1.5$

4. $f(x)=\sin x,\ x_0=1$

5. $f(x)=x^2+x-2,\ x_0=0$

6. $f(x)=x^2-2,\ x_0=1.5$

7. $f(x)=\ln x,\ x_0=2$

In Exercises 8-11, use Newton's Method to approximate all roots of the given functions accurate to 3 places after the decimal. If an interval is given, find only the roots that lie in that interval. Use technology to obtain good initial approximations.

8. $f(x)=x^3+5x^2-x-1$

9. $f(x)=x^4+2x^3-7x^2-x+5$

10. $f(x)=x^{17}-2x^{13}-10x^8+10 \text{ on } (-2,2)$

11. $f(x)=x^2\cos x+(x-1)\sin x \text{ on } (-3,3)$

In exercises 12-15, use Newton's Method to approximate when the given functions are equal, accurate to 3 places after the decimal. Use technology to obtain good initial approximations.

12. $f(x)=x^2,\ g(x)=\cos x$

13. $f(x)=x^2-1,\ g(x)=\sin x$

14. $f(x)=e^{x^2},\ g(x)=\cos x$

15. $f(x)=x,\ g(x)=\tan x \text{ on } [-6,6]$

16. Why does Newton's Method fail in finding the root of $f(x)=x^3-3x^2+x+3 \text{ when } x_0=1$?

17. Why does Newton's Method fail in finding the root of $f(x)=-17x^4+130x^3-301x^2+156x+156 \text{ when } x_0=1$?

4.2: Related Rates

Terms and Concepts

1. T/F: Implicit differentiation is often used when solving "related rates" type problems.

2. T/F: A study of related rates is part of the standard police officer training.

Problems

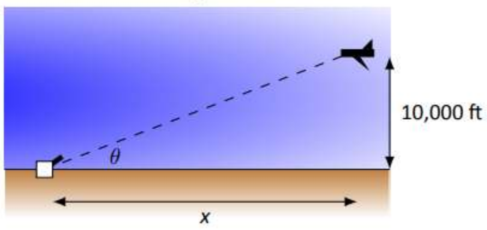
3. Water flows onto a flat surface at a rate of 5cm^3/s forming a circular puddle 10mm deep. How fast is the radius growing when the radius is:
(a) 1 cm?
(b) 10 cm?
(c) 100 cm?

4. A circular balloon is inflated with air flowing at a rate of 10cm^3/s. How fast is the radius of the balloon increasing when the radius is:
(a) 1 cm?
(b) 10 cm?
(c) 100 cm?

5. Consider the traffic situation introduced in Example 100. How fast is the “other car” traveling if the officer and the other car are each 1/2 mile from the intersection, the other car is traveling *due west*, the officer is traveling north at 50mph, and the radar reading is −80mph?

6. Consider the traffic situation introduced in Example 100. Calculate how fast the “other car” is traveling in each of the following situations.
(a) The officer is traveling due north at 50mph and is 1/2 mile from the intersection, while the other car is 1 mile from the intersection traveling west and the radar reading is −80mph?
(b) The officer is traveling due north at 50mph and is 1 mile from the intersection, while the other car is 1/2 mile from the intersection traveling west and the radar reading is −80mph?

7. An F-22 aircraft is flying at 500mph with an elevation of 10,000ft on a straight–line path that will take it directly over an anti–aircraft gun.

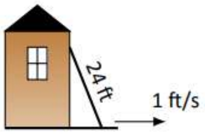


How fast must the gun be able to turn to accurately track the aircraft when the plane is:
(a) 1 mile away?
(b) 1/5 mile away?
(c) Directly overhead?

8. An F-22 aircraft is flying at 500mph with an elevation of 100ft on a straight–line path that will take it directly over an anti–aircraft gun as in Exercise 7 (note the lower elevation here).

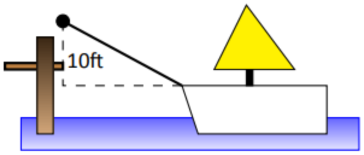
How fast must the gun be able to turn to accurately track the aircraft when the plane is:
(a) 1000 feet away?
(b) 100 feet away?
(c) Directly overhead?

9. A 24ft. ladder is leaning against a house while the base is pulled away at a constant rate of 1 ft/s.



At what rate is the top of the ladder sliding down the side of the house when the base is:
(a) 1 foot from the house?
(b) 10 feet from the house?
(c) 23 feet from the house?
(d) 24 feet from the house?

10. A boat is being pulled into a dock at a constant rate of 30ft/min by a winch located 10ft above the deck of the boat.

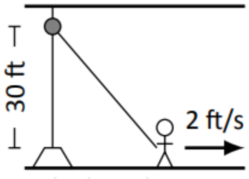


At what rate is the boat approaching the dock when the boat is:
(a) 50 feet out?
(b) 15 feet out?
(c) 1 foot from the dock?
(d) What happens when the length of rope pulling in the boat is less than 10 feet long?

11. An inverted cylindrical cone, 20ft deep and 10ft across at the top, is being filled with water at a rate of $10ft^3$/min. At what rate is the water rising in the tank when the depth of the water is:
(a) 1 foot?
(b) 10 feet?
(c) 19 feet?

How long will the tank take to fill when starting at empty?

12. A rope, attached to a weight, goes up through a pulley at the ceiling and back down to a worker. The man holds the rope at the same height as the connection point between rope and weight.



Suppose the man stands directly next to the weight (i.e., a total rope length of 60ft) and begins to walk away at a rate of 2ft/s. How fast is the weight rising when the man has walked:
(a) 10 feet?
(b) 40 feet?

How far must the man walk to raise the weight all the way to the pulley?

13. Consider the situation described in Exercise 12. Suppose the man starts 40ft from the weight and begins to walk away at a rate of 2ft/s.
(a) How long is the rope?
(b) How fast is the weight rising after the man has walked 10 feet?
(c) How fast is the weight rising after the man has walked 40 feet?
(d) How far must the man walk to raise the weight all the way to the pulley?

14. A hot air balloon lifts off from ground rising vertically. From 100 feet away, a 5' woman tracks the path of the balloon. When her sightline with the balloon makes a 45 degree angle with the horizontal, she notes the angle is increasing at about 5° /min.

15. A company that produces landscaping materials is dumping sand into a conical pile. The sand is being poured at a rate of 5ft ^{3}/sec; the physical properties of the sand, in conjunction with gravity, ensure that the cone's height is roughly 2/3 the length of the diameter of the circular base.
How fast is the cone rising when it has a height of 30 feet?

4.3: Optimization

Terms and Concepts

1. T/F: An "optimization problem" is essentially an "extreme values" problem in a "story problem" setting.

2. T/F: This section teaches one to find the extreme values of function that have more than one variable.

Problems

3. Find the maximum product of two numbers (not necessarily integers) that have a sum of 100.

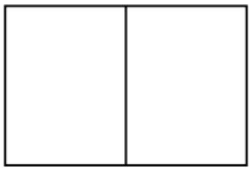
4. Find the minimum sum of two numbers whose product is 500.

5. Find the maximum sum of two numbers whose product is 500.

6. Find the maximum sum of two numbers, each of which is in [0,300] whose product is 500.

7. Find the maximal area of a right triangle with hypotenuse of length 1.

8. A rancher has 1000 feet of fencing in which to construct adjacent, equally sized rectangular pens. What dimensions should these pens have to maximize the enclosed area?



9. A standard soda can is roughly cylindrical and holds 355cm^3 of liquid. What dimensions should the cylinder be to minimize the material needed to produce the can? Based on your dimensions, determine whether or not the standard can is produced to minimize the material costs.

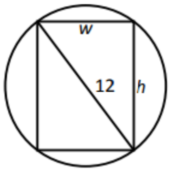
10. Find the dimensions of a cylindrical can with a volume of 206in^3 that minimizes the surface area.

The “#10 can”is a standard sized can used by the restaurant industry that holds about 206in^3 with a diameter of 6 2/16in and height of 7in. Does it seem these dimensions where chosen with minimization in mind?

11. The United States Postal Service charges more for boxes whose combined length and girth exceeds 108” (the “length” of a package is the length of its longest side; the girth is the perimeter of the cross section, i.e., 2w + 2h)

What is the maximum volume of a package with a square cross section (w = h) that does not exceed the 108” standard?

12. The strength S of a wooden beam is directly proportional to its cross sectional width w and the square of its height h; that is, $S = kwh^2$ for some constant k.



Given a circular log with diameter of 12 inches, what sized beam can be cut from the log with maximum strength?

13. A power line is to be run to an offshore facility in the manner described in Example 104. The offshore facility is 2 miles at sea and 5 miles along the shoreline from the power plant. It costs $50,000 per mile to lay a power line underground and $80,000 to run the line underwater.

How much of the power line should be run underground to minimize the overall costs?

14. A power line is to be run to an offshore facility in the manner described in Example 104. The offshore facility is 5 miles at sea and 2 miles along the shoreline from the power plant. It costs $50,000 per mile to lay a power line underground and $80,000 to run the line underwater.

How much of the power line should be run underground to minimize the overall costs?

15. A woman throws a stick into a lake for her dog to fetch; the stick is 20 feet down the shore line and 15 feet into the water from there. The dog may jump directly into the water and swim, or run along the shore line to get closer to the stick before swimming. The dog runs about 22ft/s and swims about 1.5ft/s.

How far along the shore should the dog run to minimize the time it takes to get to the stick? (Hint: the figure from Example 104 can be useful.)

16. A woman throws a stick into a lake for her dog to fetch; the stick is 15 feet down the shore line and 30 feet into the water from there. The dog may jump directly into the water and swim, or run along the shore line to get closer to the stick before swimming. The dog runs about 22ft/s and swims about 1.5ft/s.

How far along the shore should the dog run to minimize the time it takes to get to the stick? (Google "calculus dog" to learn more about a dog's ability to minimize times.)

17. What are the dimensions of the rectangle with largest area that can be drawn inside the unit circle?

4.4: Differentials

Terms and Concepts

1. T/F: Given a differentiable function $y = f(x)$, we are generally free to choose a value for dx, which then determines the value of dy.

2. T/F: the symbols $''dx''$ and $''\Delta x''$ represent the same concept.

3. T/F: the symbols $''dy''$ and $''\Delta y''$ represent the same concept.

4. T/F: Differentials are important in the study of integration.

5. How are the differentials and tangent lines related?

Problems

In Exercises 6-17, use differentials to approximate the given value by hand.

6. 2.05^2

7. 5.93^2

8. 5.1^3

9. 6.8^3

10. $\sqrt{16.5}$

11. $\sqrt{24}$

12. $\sqrt[3]{63}$

13. $\sqrt[3]{8.5}$

14. $\sin 3$

15. $\cos 1.5$

16. $e^{0.1}$

In Exercises 17-29, compute the differential dy.

17. $y = x^2 + 3x - 5$

18. $y = x^7 - x^5$

19. $y = \frac{1}{4x^2}$

20. $y = (2x + \sin x)^2$

21. $y = x^2 e^{3x}$

22. $y = \frac{4}{x^4}$

23. $y = \frac{2x}{\tan x + 1}$

24. $y = \ln(5x)$

25. $y = e^x \sin x$

26. $y = \cos(\sin x)$

27. $y = \frac{x+1}{x+2}$

28. $y = 3^x \ln x$

29. $y = x \ln x - x$

30. A set of plastic spheres are to be made with a diameter of 1cm. If the manufacturing process is accurate to 1mm, what is the propagated error in volume of the spheres?

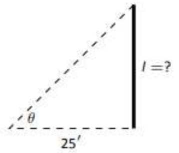
31. The distance, in feet, a stone drops in t seconds is given by $d(t) = 16t^2$. The depth of a hole is to be approximated by dropping a rock and listening for it to hit the bottom. What is the propagated error if the time measurement is accurate to 2/10ths of a second and the measured time is:
(a) 2 seconds?
(b) 5 seconds?

32. What is the propagated error in the measurement of the cross sectional area of a circular log if the diameter is measured at 15″, accurate to 1/4 ″?

33. A wall is to be painted that is 8′ high and is measured to be 10′ , 7″ long. Find the propagated error in the measurement of the wall's surface area if the measurement is accurate to 1/2 ″ .

Exercises 34-38 explore some issues related to surveying in which distances are approximated using other measured distance and measured angles. (Hint: Convert all angle to radians before computing.)

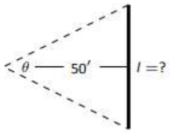
34. The length l of a long wall is to be approximated. The angle θ, as shown in the diagram (not to scale), is measured to be 85.2°, accurate to 1°. Assume that the triangle formed is a right triangle.



(a) What is the measured length l of the wall?
(b) What is the propagated error?
(c) What is the percent error?

35. Answer the questions to Exercise 34, but with a measured angle of 71.5°, accurate to 1°, measured with a point 100' from the wall.

36. The length l of a long wall is to be calculated by measuring the angle θ shown in the diagram (not to scale). Assume the formed triangle is an isosceles triangle. The measured angle is 143°, accurate to 1°.



(a) What is the measured length of the wall?
(b) what is the propagated error?
(c) What is the percent error?

37. The length of the walls in Exercise 34-36 are essentially the same. Which setup gives the most accurate result?

38. consider the setup in Exercises 36. This time, assume the angle measurement of 143° is exact but the measured 50° from the wall is accurate to 6″. What is the approximate percent error?

4.E: Applications of Derivatives (Exercises) is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

5: Integration

5.1: Antiderivatives and Indefinite Integration
5.2: The Definite Integral
5.3: Riemann Sums
5.4: The Fundamental Theorem of Calculus
5.5: Numerical Integration
5.E: Applications of Integration (Exercises)

Contributors

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 5: Integration is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

5.1: Antiderivatives and Indefinite Integration

Antiderivatives

We have spent considerable time considering the derivatives of a function and their applications. In the following chapters, we are going to starting thinking in "the other direction." That is, given a function $f(x)$, we are going to consider functions $F(x)$ such that $F'(x) = f(x)$. There are numerous reasons this will prove to be useful: these functions will help us compute areas, volumes, mass, force, pressure, work, and much more.

Given a function $y = f(x)$, a differential equation is one that incorporates y, x, and the derivatives of y. For instance, a simple differential equation is:

$$y' = 2x. \tag{5.1.1}$$

Solving a differential equation amounts to finding a function y that satisfies the given equation. Take a moment and consider that equation; can you find a function y such that $y' = 2x$?

Can you find another?

And yet another?

Hopefully one was able to come up with at least one solution: $y = x^2$. "Finding another" may have seemed impossible until one realizes that a function like $y = x^2 + 1$ also has a derivative of $2x$. Once that discovery is made, finding "yet another" is not difficult; the function $y = x^2 + 123,456,789$ also has a derivative of $2x$. The differential equation $y' = 2x$ has many solutions. This leads us to some definitions.

Definition5.1.1: Antiderivatives and Indefinite Integrals

Let a function $f(x)$ be given. An **antiderivative** of $f(x)$ is a function $F(x)$ such that $F'(x) = f(x)$.

The set of all antiderivatives of $f(x)$ is the **indefinite integral of** f, denoted by

$$\int f(x)\ dx. \tag{5.1.2}$$

Make a note about our definition: we refer to *an* antiderivative of f, as opposed to *the* antiderivative of f, since there is *always* an infinite number of them. We often use upper-case letters to denote antiderivatives.

Knowing one antiderivative of f allows us to find infinitely more, simply by adding a constant. Not only does this give us *more* antiderivatives, it gives us *all* of them.

Theorem 5.1.1: Antiderivative Forms

Let $F(x)$ and $G(x)$ be antiderivatives of $f(x)$. Then there exists a constant C such that

$$G(x) = F(x) + C. \tag{5.1.3}$$

Given a function f and one of its antiderivatives F, we know *all* antiderivatives of f have the form $F(x) + C$ for some constant C. Using Definition $5.1.1$, we can say that

$$\int f(x)\ dx = F(x) + C. \tag{5.1.4}$$

Let's analyze this indefinite integral notation.

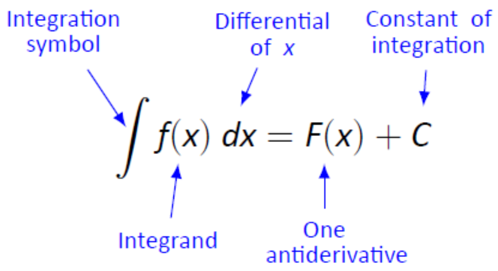


***Figure** 5.1.1: Understanding the indefinite integral notation.*

Figure 5.1.1 shows the typical notation of the indefinite integral. The integration symbol, $\int$, is in reality an "elongated S," representing "take the sum." We will later see how *sums* and *antiderivatives* are related.

The function we want to find an antiderivative of is called the *integrand*. It contains the differential of the variable we are integrating with respect to. The $\int$ symbol and the differential dx are not "bookends" with a function sandwiched in between; rather, the symbol $\int$ means "find all antiderivatives of what follows," and the function $f(x)$ and dx are multiplied together; the dx does not "just sit there."

Let's practice using this notation.

Example 5.1.1: Evaluating indefinite integrals

Evaluate $\int \sin x\ dx$.

Solution

We are asked to find all functions $F(x)$ such that $F'(x) = \sin x$. Some thought will lead us to one solution: $F(x) = -\cos x$, because $\frac{d}{dx}(-\cos x) = \sin x$.

The indefinite integral of $\sin x$ is thus $-\cos x$, plus a constant of integration. So:

$$\int \sin x\ dx = -\cos x + C. \tag{5.1.5}$$

A commonly asked question is "What happened to the dx?" The unenlightened response is "Don't worry about it. It just goes away." A full understanding includes the following.

This process of *antidifferentiation* is really solving a *differential* question. The integral

$$\int \sin x\ dx \tag{5.1.6}$$

presents us with a differential, $dy = \sin x\ dx$. It is asking: "What is y?" We found lots of solutions, all of the form $y = -\cos x + C$.

Letting $dy = \sin x\ dx$, rewrite

$$\int \sin x\ dx \quad \text{as} \quad \int dy. \tag{5.1.7}$$

This is asking: "What functions have a differential of the form dy?" The answer is "Functions of the form $y + C$, where C is a constant." What is y? We have lots of choices, all differing by a constant; the simplest choice is $y = -\cos x$.

Understanding all of this is more important later as we try to find antiderivatives of more complicated functions. In this section, we will simply explore the rules of indefinite integration, and one can succeed for now with answering "What happened to the dx?" with "It went away."

Let's practice once more before stating integration rules.

Example 5.1.2: Evaluating indefinite integrals

Evaluate $\int(3x^2+4x+5)\ dx$.

Solution

We seek a function $F(x)$ whose derivative is $3x^2+4x+5$. When taking derivatives, we can consider functions term--by--term, so we can likely do that here.

What functions have a derivative of $3x^2$? Some thought will lead us to a cubic, specifically x^3+C_1, where C_1 is a constant.

What functions have a derivative of $4x$? Here the x term is raised to the first power, so we likely seek a quadratic. Some thought should lead us to $2x^2+C_2$, where C_2 is a constant.

Finally, what functions have a derivative of 5? Functions of the form $5x+C_3$, where C_3 is a constant.

Our answer appears to be

$$\int(3x^2+4x+5)\ dx=x^3+C_1+2x^2+C_2+5x+C_3. \tag{5.1.8}$$

We do not need three separate constants of integration; combine them as one constant, giving the final answer of

$$\int(3x^2+4x+5)\ dx=x^3+2x^2+5x+C. \tag{5.1.9}$$

It is easy to verify our answer; take the derivative of x^3+2x^3+5x+C and see we indeed get $3x^2+4x+5$.

This final step of "verifying our answer" is important both practically and theoretically. In general, taking derivatives is easier than finding antiderivatives so checking our work is easy and vital as we learn.

We also see that taking the derivative of our answer returns the function in the integrand. Thus we can say that:

$$\frac{d}{dx}\left(\int f(x)\ dx\right)=f(x). \tag{5.1.10}$$

Differentiation "undoes" the work done by antidifferentiation.

Theorem 27 gave a list of the derivatives of common functions we had learned at that point. We restate part of that list here to stress the relationship between derivatives and antiderivatives. This list will also be useful as a glossary of common antiderivatives as we learn.

Theorem 5.1.2: Derivatives and Antiderivatives

Common Differentiation Rules	Common Indefinite Integration Rules
1. $\frac{d}{dx}\big(cf(x)\big)=c\cdot f'(x)$	1. $\int c\cdot f(x)\ dx=c\cdot\int f(x)\ dx$
2. $\frac{d}{dx}\big(f(x)\pm g(x)\big)=f'(x)\pm g'(x)$	2. $\int\big(f(x)\pm g(x)\big)\ dx=\int f(x)\ dx\pm\int g(x)\ dx$
3. $\frac{d}{dx}\big(C\big)=0$	3. $\int 0\ dx=C$
4. $\frac{d}{dx}\big(x\big)=1$	4. $\int 1\ dx=\int dx=x+C$
5. $\frac{d}{dx}\big(x^n\big)=n\cdot x^{n-1}$	5. $\int x^n\ dx=\frac{1}{n+1}x^{n+1}+C$
6. $\frac{d}{dx}\big(\sin x\big)=\cos x$	6. $\int\cos x\ dx=\sin x+C$
7. $\frac{d}{dx}\big(\cos x\big)=-\sin x$	7. $\int\sin x\ dx=-\cos x+C$
8. $\frac{d}{dx}\big(\tan x\big)=\sec^2 x$	8. $\int\sec^2 x\ dx=\tan x+C$
9. $\frac{d}{dx}\big(\csc x\big)=-\csc x\cot x$	9. $\int\csc x\cot x\ dx=-\csc x+C$
10. $\frac{d}{dx}\big(\sec x\big)=\sec x\tan x$	10. $\int\sec x\tan x\ dx=\sec x+C$
11. $\frac{d}{dx}\big(\cot x\big)=-\csc^2 x$	11. $\int\csc^2 x\ dx=-\cot x+C$
12. $\frac{d}{dx}\big(e^x\big)=e^x$	12. $\int e^x\ dx=e^x+C$
13. $\frac{d}{dx}\big(a^x\big)=\ln a\cdot a^x$	13. $\int a^x\ dx=\frac{1}{\ln a}\cdot a^x+C$
14. $\frac{d}{dx}\big(\ln x\big)=\frac{1}{x}$	14. $\int\frac{1}{x}\ dx=\ln\lvert x\rvert+C$

We highlight a few important points from Theorem $5.1.2$:

- Rule #1 states $\int c\cdot f(x)\ dx=c\cdot \int f(x)\ dx$. This is the Constant Multiple Rule: we can temporarily ignore constants when finding antiderivatives, just as we did when computing derivatives (i.e., $\frac{d}{dx}\left(3x^2\right)$ is just as easy to compute as $\frac{d}{dx}\left(x^2\right)$). An example:

$$\int 5\cos x\ dx=5\cdot \int \cos x\ dx=5\cdot (\sin x+C)=5\sin x+C. \tag{5.1.11}$$

In the last step we can consider the constant as also being multiplied by 5, but "5 times a constant" is still a constant, so we just write "C,".

- Rule #2 is the Sum/Difference Rule: we can split integrals apart when the integrand contains terms that are added/subtracted, as we did in Example 5.1.2. So:

$$\begin{align}\int (3x^2+4x+5)\ dx &= \int 3x^2\ dx+\int 4x\ dx+\int 5\ dx \tag{5.1.12}\\ &=3\int x^2\ dx+4\int x\ dx+\int 5\ dx \tag{5.1.13}\\ &=3\cdot \frac{1}{3}x^3+4\cdot \frac{1}{2}x^2+5x+C \tag{5.1.14}\\ &=x^3+2x^2+5x+C \tag{5.1.15}\end{align}$$

In practice we generally do not write out all these steps, but we demonstrate them here for completeness.

- Rule #5 is the Power Rule of indefinite integration. There are two important things to keep in mind:
  1. Notice the restriction that $n\neq -1$. This is important: $\int \frac{1}{x}\ dx\neq$ "$\frac{1}{0}x^0+C$"; rather, see Rule #14.
  2. We are presenting antidifferentiation as the "inverse operation" of differentiation. Here is a useful quote to remember: "Inverse operations do the opposite things in the opposite order."
     When taking a derivative using the Power Rule, we **first** *multiply* by the power, then **second** *subtract* 1 from the power. To find the antiderivative, do the opposite things in the opposite order: **first** *add* one to the power, then **second** *divide* by the power.
- Note that Rule #14 incorporates the absolute value of x. The exercises will work the reader through why this is the case; for now, know the absolute value is important and cannot be ignored.

Initial Value Problems

In Section 2.3 we saw that the derivative of a position function gave a velocity function, and the derivative of a velocity function describes acceleration. We can now go "the other way:" the antiderivative of an acceleration function gives a velocity function, etc. While there is just one derivative of a given function, there are infinite antiderivatives. Therefore we cannot ask "What is *the* velocity of an object whose acceleration is –32ft/s^2?", since there is more than one answer.

We can find *the* answer if we provide more information with the question, as done in the following example. Often the additional information comes in the form of an *initial value*, a value of the function that one knows beforehand.

Example 5.1.3: Solving initial value problems

The acceleration due to gravity of a falling object is -32 ft/s^2. At time $t=3$, a falling object had a velocity of -10 ft/s. Find the equation of the object's velocity.

Solution

We want to know a velocity function, $v(t)$. We know two things:

1. The acceleration, i.e., $v'(t)=-32$, and
2. the velocity at a specific time, i.e., $v(3)=-10$.

Using the first piece of information, we know that $v(t)$ is an antiderivative of $v'(t)=-32$. So we begin by finding the indefinite integral of -32:

$$\int (-32)\ dt=-32t+C=v(t). \tag{5.1.16}$$

Now we use the fact that $v(3)=-10$ to find C:

$$v(t) = -32t + C \tag{5.1.17}$$
$$v(3) = -10 \tag{5.1.18}$$
$$-32(3) + C = -10 \tag{5.1.19}$$
$$C = 86 \tag{5.1.20}$$

Thus $v(t) = -32t + 86$. We can use this equation to understand the motion of the object: when $t = 0$, the object had a velocity of $v(0) = 86$ ft/s. Since the velocity is positive, the object was moving upward.

When did the object begin moving down? Immediately after $v(t) = 0$:

$$-32t + 86 = 0 \quad \Rightarrow \quad t = \frac{43}{16} \approx 2.69\mathrm{s}. \tag{5.1.21}$$

Recognize that we are able to determine quite a bit about the path of the object knowing just its acceleration and its velocity at a single point in time.

Example $5.1.4$: Solving initial value problems

Find $f(t)$, given that $f''(t) = \cos t$, $f'(0) = 3$ and $f(0) = 5$.

Solution

We start by finding $f'(t)$, which is an antiderivative of $f''(t)$:

$$\int f''(t)\ dt = \int \cos t\ dt = \sin t + C = f'(t). \tag{5.1.22}$$

So $f'(t) = \sin t + C$ for the correct value of C. We are given that $f'(0) = 3$, so:

$$f'(0) = 3 \quad \Rightarrow \quad \sin 0 + C = 3 \quad \Rightarrow \quad C = 3. \tag{5.1.23}$$

Using the initial value, we have found $f'(t) = \sin t + 3$.

We now find $f(t)$ by integrating again.

$$f(t) = \int f'(t)\ dt = \int (\sin t + 3)\ dt = -\cos t + 3t + C. \tag{5.1.24}$$

We are given that $f(0) = 5$, so

$$-\cos 0 + 3(0) + C = 5 \tag{5.1.25}$$
$$-1 + C = 5 \tag{5.1.26}$$
$$C = 6 \tag{5.1.27}$$

Thus $f(t) = -\cos t + 3t + 6$.

This section introduced antiderivatives and the indefinite integral. We found they are needed when finding a function given information about its derivative(s). For instance, we found a position function given a velocity function.

In the next section, we will see how position and velocity are unexpectedly related by the areas of certain regions on a graph of the velocity function. Then, in Section 5.4, we will see how areas and antiderivatives are closely tied together.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/
- Integrated by Justin Marshall.

This page titled 5.1: Antiderivatives and Indefinite Integration is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

5.2: The Definite Integral

We start with an easy problem. An object travels in a straight line at a constant velocity of 5 ft/s for 10 seconds. How far away from its starting point is the object?

We approach this problem with the familiar "Distance $=$ Rate $\times$ Time" equation. In this case, Distance = 5ft/s $\times$ 10s $=$ 50 feet.

It is interesting to note that this solution of 50 feet can be represented graphically. Consider Figure $5.2.1$, where the constant velocity of 5ft/s is graphed on the axes. Shading the area under the line from $t=0$ to $t=10$ gives a rectangle with an area of 50 square units; when one considers the units of the axes, we can say this area represents 50 ft.

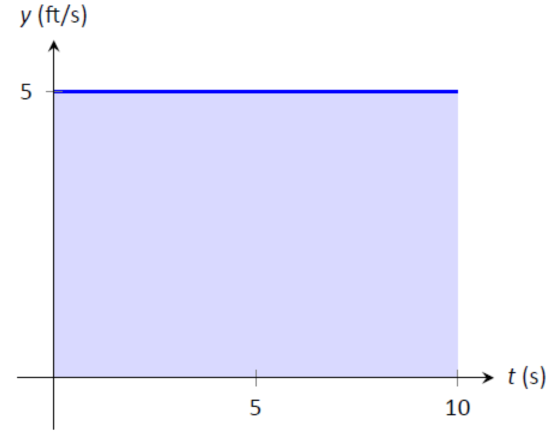


***Figure** $5.2.1$**:** The area under a constant velocity function corresponds to distance traveled.*

Now consider a slightly harder situation (and not particularly realistic): an object travels in a straight line with a constant velocity of 5ft/s for 10 seconds, then instantly reverses course at a rate of 2ft/s for 4 seconds. (Since the object is traveling in the opposite direction when reversing course, we say the velocity is a constant -2ft/s.) How far away from the starting point is the object -- what is its *displacement*?

Here we use "Distance $=$ Rate$_1$ $\times$ Time$_1$ + Rate$_2$ $\times$ Time$_2$," which is

$$\text{Distance} \quad = 5\cdot 10+(-2)\cdot 4 = 42 \text{ ft}. \tag{5.2.1}$$

Hence the object is 42 feet from its starting location.

We can again depict this situation graphically. In Figure $5.2.3$ we have the velocities graphed as straight lines on $[0,10]$ and $[10,14]$, respectively. The displacement of the object is

$$\text{"Area above the } t\text{--axis} \quad - \quad \text{Area below the } t\text{-axis}, \tag{5.2.2}$$

which is easy to calculate as $50-8=42$ feet.

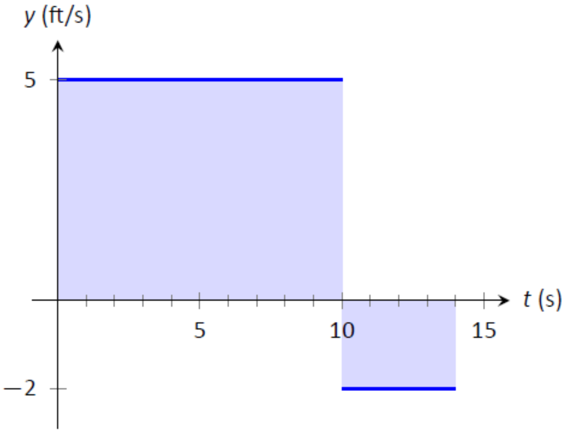


***Figure** $5.2.2$: The total displacement is the area above the t--axis minus the area below the t--axis.*

Now consider a more difficult problem.

Example $5.2.1$: Finding position using velocity

The velocity of an object moving straight up/down under the acceleration of gravity is given as $v(t) = -32t + 48$, where time t is given in seconds and velocity is in ft/s. When $t = 0$, the object had a height of 0 ft.

1. What was the initial velocity of the object?
2. What was the maximum height of the object?
3. What was the height of the object at time $t = 2$?

Solution

It is straightforward to find the initial velocity; at time $t = 0$, $v(0) = -32 \cdot 0 + 48 = 48$ ft/s.

To answer questions about the height of the object, we need to find the object's position function $s(t)$. This is an initial value problem, which we studied in the previous section. We are told the initial height is 0, i.e., $s(0) = 0$. We know $s'(t) = v(t) = -32t + 48$. To find s, we find the indefinite integral of $v(t)$:

$$\int v(t)\ dt = \int (-32t + 48)\ dt = -16t^2 + 48t + C = s(t). \tag{5.2.3}$$

Since $s(0) = 0$, we conclude that $C = 0$ and $s(t) = -16t^2 + 48t$.

To find the maximum height of the object, we need to find the maximum of s. Recalling our work finding extreme values, we find the critical points of s by setting its derivative equal to 0 and solving for t:

$$s'(t) = -32t + 48 = 0 \quad \Rightarrow \quad t = 48/32 = 1.5\text{s}. \tag{5.2.4}$$

(Notice how we ended up just finding when the velocity was 0ft/s!) The first derivative test shows this is a maximum, so the maximum height of the object is found at

$$s(1.5) = -16(1.5)^2 + 48(1.5) = 36\text{ft}. \tag{5.2.5}$$

The height at time $t = 2$ is now straightforward to compute: it is $s(2) = 32$ft.

While we have answered all three questions, let's look at them again graphically, using the concepts of area that we explored earlier.

Figure $5.2.3$ shows a graph of $v(t)$ on axes from $t = 0$ to $t = 3$. It is again straightforward to find $v(0)$. How can we use the graph to find the maximum height of the object?

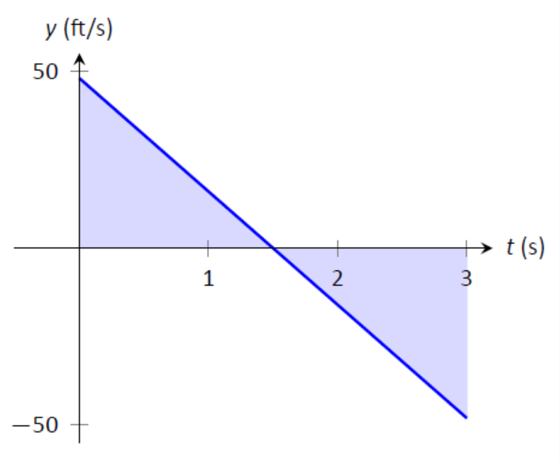


***Figure** $5.2.3$:* A graph of $v(t) = -32t + 48$; the shaded areas help determine displacement.

Recall how in our previous work that the displacement of the object (in this case, its height) was found as the area under the velocity curve, as shaded in the figure. Moreover, the area between the curve and the t--axis that is below the t--axis counted as "negative" area. That is, it represents the object coming back toward its starting position. So to find the maximum distance from the starting point -- the maximum height -- we find the area under the velocity line that is above the t--axis, i.e., from $t = 0$ to $t = 1.5$. This region is a triangle; its area is

$$\text{Area } = \frac{1}{2}\text{Base} \times \text{Height} = \frac{1}{2} \times 1.5\text{s} \times 48\text{ft/s} = 36\text{ft}, \tag{5.2.6}$$

which matches our previous calculation of the maximum height.

Finally, we find the total *signed* area under the velocity function from $t = 0$ to $t = 2$ to find the $s(2)$, the height at $t = 2$, which is a displacement, the distance from the current position to the starting position. That is,

$$\text{Displacement} = \text{Area above the } t\text{—axis} - \text{Area below } t\text{-axis}. \tag{5.2.7}$$

The regions are triangles, and we find

$$\text{Displacement} = \frac12(1.5\text{s})(48\text{ft/s}) - \frac12(.5\text{s})(16\text{ft/s}) = 32\text{ft}.\]

This also matches our previous calculation of the height at $t = 2$.

Notice how we answered each question in this example in two ways. Our first method was to manipulate equations using our understanding of antiderivatives and derivatives. Our second method was geometric: we answered questions looking at a graph and finding the areas of certain regions of this graph.

The above example does not *prove* a relationship between area under a velocity function and displacement, but it does imply a relationship exists. Section 5.4 will fully establish fact that the area under a velocity function is displacement.

Given a graph of a function $y = f(x)$, we will find that there is great use in computing the area between the curve $y = f(x)$ and the x-axis. Because of this, we need to define some terms.

Definition 5.2.1: The Definite Integral, Total Signed Area

Let $y = f(x)$ be defined on a closed interval $[a, b]$. The **total signed area from $x = a$ to $x = b$ under f** is:

$$\begin{gathered}(\text{area under } f \text{ and above the } x\text{-axis on } [a, b]) \\ . - (\text{area above } f \text{ and under the } x\text{-axis on } [a, b]).\end{gathered} \tag{5.2.8}$$

The **definite integral of f on** $[a, b]$ is the total signed area of f on $[a, b]$, denoted

$$\int_a^b f(x)\ dx, \tag{5.2.9}$$

There a and b are the **bounds of integration.**

By our definition, the definite integral gives the "signed area under f." We usually drop the word "signed" when talking about the definite integral, and simply say the definite integral gives "the area under f\," or, more commonly, "the area under the curve."

The previous section introduced the indefinite integral, which related to antiderivatives. We have now defined the definite integral, which relates to areas under a function. The two are very much related, as we'll see when we learn the Fundamental Theorem of Calculus in Section 5.4. Recall that earlier we said that the "$\int$" symbol was an "elongated S" that represented finding a "sum." In the context of the definite integral, this notation makes a bit more sense, as we are adding up areas under the function f.

We practice using this notation.

Example 5.2.2: Evaluating definite integrals

Consider the function f given in Figure 5.2.4.

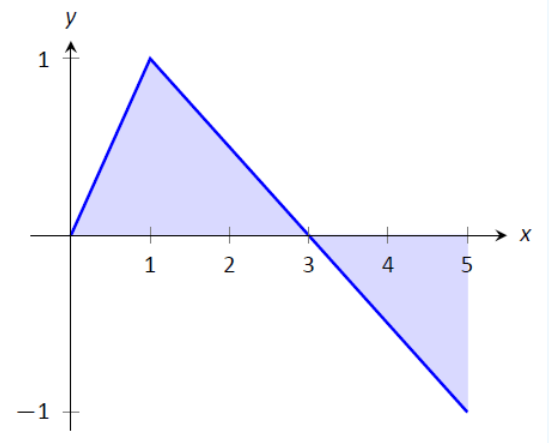


Figure $5.2.4$***:*** A graph of $f(x)$ in Example 2

Find:

1. $\int_0^3 f(x)\ dx$
2. $\int_3^5 f(x)\ dx$
3. $\int_0^5 f(x)\ dx$
4. $\int_0^3 5f(x)\ dx$
5. $\int_1^1 f(x)\ dx$

Solution

1. $\int_0^3 f(x)dx$ is the area under f on the interval $[0,3]$. This region is a triangle, so the area is $\int_0^3 f(x)dx = \frac{1}{2}(3)(1) = 1.5$.
2. $\int_3^5 f(x)dx$ represents the area of the triangle found under the x--axis on $[3,5]$. The area is $\frac{1}{2}(2)(1) = 1$; since it is found *under* the x--axis, this is "negative area." Therefore $\int_3^5 f(x)\ dx = -1$.
3. $\int_0^5 f(x)dx$ is the total signed area under f on $[0,5]$. This is $1.5 + (-1) = 0.5$.
4. $\int_0^3 5f(x)dx$ is the area under $5f$ on $[0,3]$. This is sketched in Figure $5.2.5$. Again, the region is a triangle, with height 5 times that of the height of the original triangle. Thus the area is $\int_0^3 5f(x)\ dx = 15/2 = 7.5$.

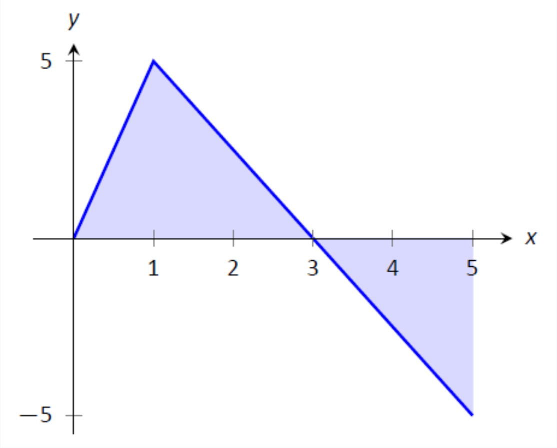


Figure $5.2.5$***:*** *A graph of* $5f$ *in Example* $5.2.2$*. (Yes, it looks just like the graph of* f *in Figure* $5.2.4$*, just with a different* y*-scale.)*

5. $\int_1^1 f(x)dx$ is the area under f on the "interval" $[1,1]$. This describes a line segment, not a region; it has no width. Therefore the area is 0.

This example illustrates some of the properties of the definite integral, given here.

Theorem $5.2.2$: Properties of the Definite Integral

Let f and g be defined on a closed interval I that contains the values a, b and c, and let k be a constant. The following hold:

1. $\int_a^a f(x)dx = 0$
2. $\int_a^b f(x)dx + \int_b^c f(x)dx = \int_a^c f(x)dx$
3. $\int_a^b f(x)dx = -\int_b^a f(x)dx$
4. $\int_a^b \big(f(x) \pm g(x)\big)dx = \int_a^b f(x)\ dx \pm \int_a^b g(x)dx$
5. $\int_a^b k \cdot f(x)dx = k \cdot \int_a^b f(x)dx$

We give a brief justification of Theorem $5.2.2$ here.

1. As demonstrated in Example $5.2.2$, there is no "area under the curve" when the region has no width; hence this definite integral is 0.
2. This states that total area is the sum of the areas of subregions. It is easily considered when we let $a < b < c$. We can break the interval $[a, c]$ into two subintervals, $[a, b]$ and $[b, c]$. The total area over $[a, c]$ is the area over $[a, b]$ plus the area over $[b, c]$.
It is important to note that this still holds true even if $a < b < c$ is not true. We discuss this in the next point.
3. This property can be viewed a merely a convention to make other properties work well. (Later we will see how this property has a justification all its own, not necessarily in support of other properties.) Suppose $b < a < c$. The discussion from the previous point clearly justifies

$$\int_b^a f(x)dx + \int_a^c f(x)\ dx = \int_b^c f(x)dx. \tag{5.2.10}$$

However, we still claim that, as originally stated,

$$\int_a^b f(x)\ dx + \int_b^c f(x)\ dx = \int_a^c f(x)\ dx. \tag{5.2.11}$$

How do Equations $5.2.11$ and $5.2.12$ relate? Start with Equation $5.2.11$:

$$\int_b^a f(x)dx + \int_a^c f(x)dx = \int_b^c f(x)dx \tag{5.2.12}$$

$$\int_a^c f(x)\ dx = -\int_b^a f(x)dx + \int_b^c f(x)dx \tag{5.2.13}$$

Property (3) justifies changing the sign and switching the bounds of integration on the $-\int_b^a f(x)dx$ term; when this is done, Equations $5.2.11$ and $5.2.12$ are equivalent.
The conclusion is this: by adopting the convention of Property (3), Property (2) holds no matter the order of a, b and c. Again, in the next section we will see another justification for this property.
4. ,5. Each of these may be non-intuitive. Property (5) states that when one scales a function by, for instance, 7, the area of the enclosed region also is scaled by a factor of 7. Both Properties (4) and (5) can be proved using geometry. The details are not complicated but are not discussed here.

Example $5.2.3$: Evaluating definite integrals using Theorem $5.2.2$.

Consider the graph of a function $f(x)$ shown in Figure $5.2.6$.

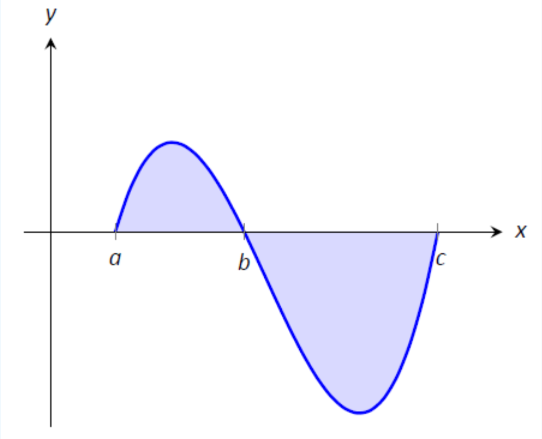


Figure $5.2.6$***:*** *A graph of a function in Example* $5.2.3$

Answer the following:

1. Which value is greater: $\int_a^b f(x)dx$ or $\int_b^c f(x)dx$?
2. Is $\int_a^c f(x)dx$ greater or less than 0?
3. Which value is greater: $\int_a^b f(x)dx$ or $\int_c^b f(x)dx$?

Solution

1. $\int_a^b f(x)dx$ has a positive value (since the area is above the x--axis) whereas $\int_b^c f(x)dx$ has a negative value. Hence $\int_a^b f(x)dx$ is bigger.
2. $\int_a^c f(x)dx$ is the total signed area under f between $x = a$ and $x = c$. Since the region below the x--axis looks to be larger than the region above, we conclude that the definite integral has a value less than 0.
3. Note how the second integral has the bounds "reversed." Therefore $\int_c^b f(x)dx$ represents a positive number, greater than the area described by the first definite integral. Hence $\int_c^b f(x)dx$ is greater.

The area definition of the definite integral allows us to use geometry compute the definite integral of some simple functions.

Example $5.2.4$: Evaluating definite integrals using geometry

Evaluate the following definite integrals:

$$1.\ \int_{-2}^{5}(2x-4)\ dx \qquad 2.\ \int_{-3}^{3}\sqrt{9-x^2}\ dx. \tag{5.2.14}$$

Solution

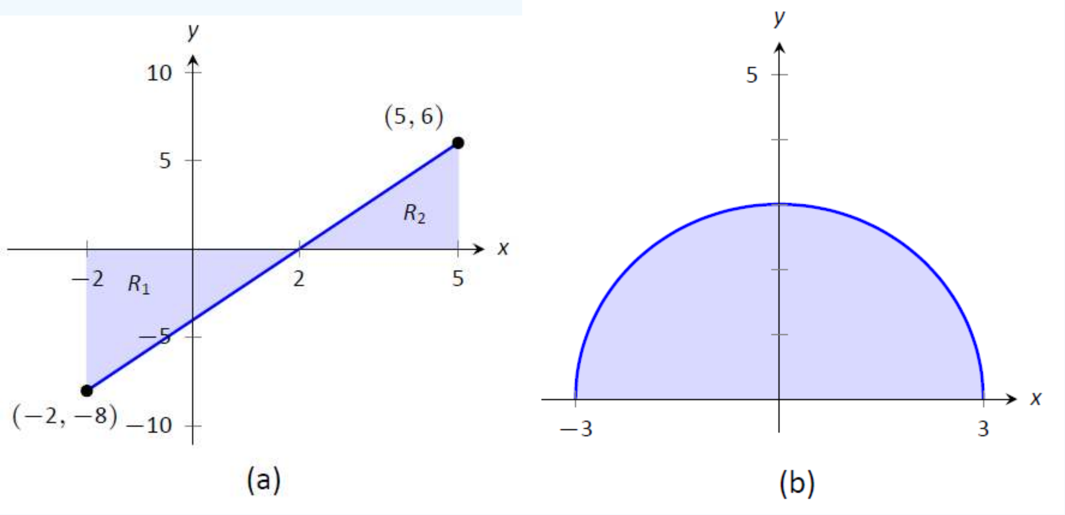


***Figure** $5.2.7$: A graph of $f(x)=2x-4$ in (a) and $f(x)=\sqrt{9-x^2}$ in (b), from Example $5.2.4$*

1. It is useful to sketch the function in the integrand, as shown in Figure $5.2.7a$. We see we need to compute the areas of two regions, which we have labeled R_1 and R_2. Both are triangles, so the area computation is straightforward:

$$R_1: \frac{1}{2}(4)(8)=16 \qquad R_2: \frac{1}{2}(3)6=9. \tag{5.2.15}$$

Region R_1 lies under the x--axis, hence it is counted as negative area (we can think of the triangle's height as being "-8"), so

$$\int_{-2}^{5}(2x-4)\ dx=-16+9=-7. \tag{5.2.16}$$

2. Recognize that the integrand of this definite integral describes a half circle, as sketched in Figure $5.2.7b$, with radius 3. Thus the area is:

$$\int_{-3}^3 \sqrt{9-x^2}\ dx = \frac12\pi r^2 = \frac 92\pi.\]

Example $5.2.5$: Understanding motion given velocity

Consider the graph of a velocity function of an object moving in a straight line, given in Figure $5.2.8$, where the numbers in the given regions gives the area of that region. Assume that the definite integral of a velocity function gives displacement. Find the maximum speed of the object and its maximum displacement from its starting position.

Solution

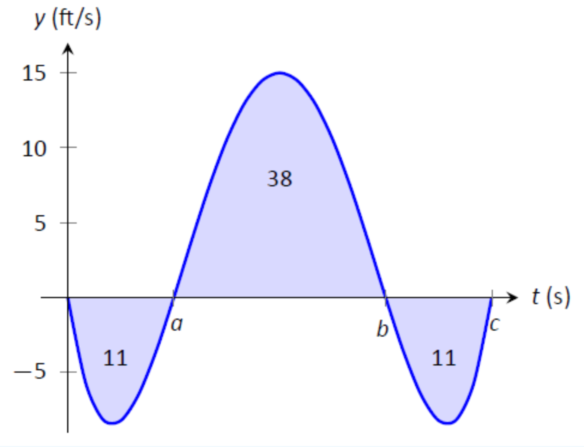


***Figure** $5.2.8$: A graph of a velocity in Example $5.2.5$.*

Since the graph gives velocity, finding the maximum speed is simple: it looks to be 15ft/s.

At time $t=0$, the displacement is 0; the object is at its starting position. At time $t=a$, the object has moved backward 11 feet. Between times $t=a$ and $t=b$, the object moves forward 38 feet, bringing it into a position 27 feet forward of its starting position. From $t=b$ to $t=c$ the object is moving backwards again, hence its maximum displacement is 27 feet from its starting position.

In our examples, we have either found the areas of regions that have nice geometric shapes (such as rectangles, triangles and circles) or the areas were given to us. Consider Figure $5.2.9$, where a region below $y=x^2$ is shaded. What is its area? The function $y=x^2$ is relatively simple, yet the shape it defines has an area that is not simple to find geometrically.

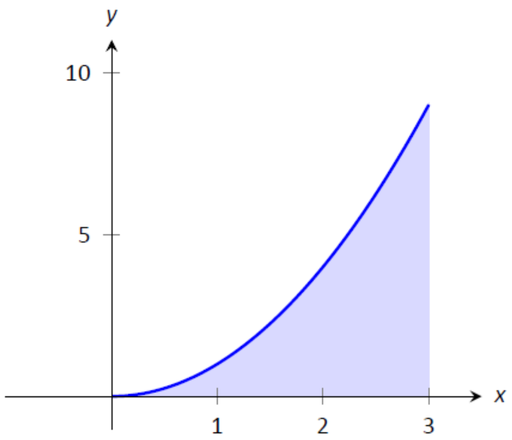


***Figure** $5.2.9$: What is the area below $y=x^2$ on $[0,3]$? The region is not a usual geometric shape.*

In the next section we will explore how to find the areas of such regions.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/
- Integrated by Justin Marshall.

This page titled 5.2: The Definite Integral is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

5.3: Riemann Sums

In the previous section we defined the definite integral of a function on $[a, b]$ to be the signed area between the curve and the x--axis. Some areas were simple to compute; we ended the section with a region whose area was not simple to compute. In this section we develop a technique to find such areas.

A fundamental calculus technique is to first answer a given problem with an approximation, then refine that approximation to make it better, then use limits in the refining process to find the exact answer. That is exactly what we will do here.

Consider the region given in Figure $5.3.1$, which is the area under $y=4x-x^2$ on $[0,4]$. What is the signed area of this region -- i.e., what is $\int_0^4 (4x-x^2)dx$?

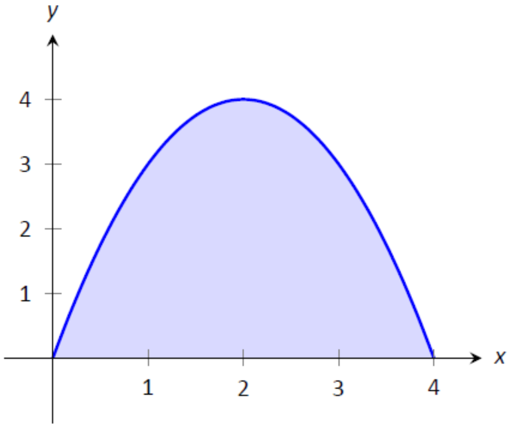


***Figure** $5.3.1$**:** A graph of $f(x)=4x-x^2$. What is the area of the shaded region?

We start by approximating. We can surround the region with a rectangle with height and width of 4 and find the area is approximately 16 square units. This is obviously an *over-approximation*; we are including area in the rectangle that is not under the parabola.

We have an approximation of the area, using one rectangle. How can we refine our approximation to make it better? The key to this section is this answer: *use more rectangles.*

Let's use 4 rectangles of equal width of 1. This *partitions* the interval $[0,4]$ into 4 *subintervals*, $[0,1]$, $[1,2]$, $[2,3]$ and $[3,4]$. On each subinterval we will draw a rectangle.

There are three common ways to determine the height of these rectangles: the **Left Hand Rule**, the **Right Hand Rule**, and the **Midpoint Rule**. The **Left Hand Rule** says to evaluate the function at the left--hand endpoint of the subinterval and make the rectangle that height. In Figure $5.3.2$, the rectangle drawn on the interval $[2,3]$ has height determined by the Left Hand Rule; it has a height of $f(2)$. (The rectangle is labeled "LHR.")

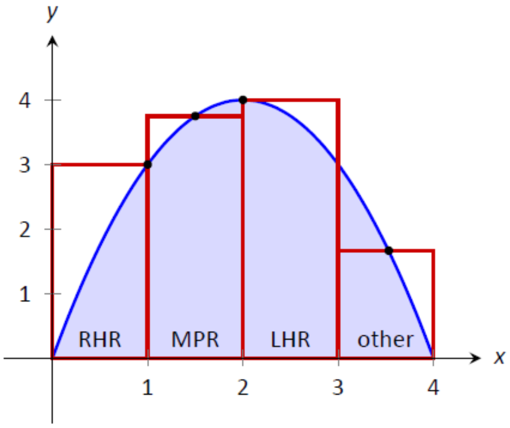


***Figure** $5.3.2$**:* Approximating $\int_0^4 (4x-x^2)dx$ using rectangles. The heights of the rectangles are determined using different rules.

The **Right Hand Rule** says the opposite: on each subinterval, evaluate the function at the right endpoint and make the rectangle that height. In the figure, the rectangle drawn on $[0,1]$ is drawn using $f(1)$ as its height; this rectangle is labeled "RHR.".

The **Midpoint Rule** says that on each subinterval, evaluate the function at the midpoint and make the rectangle that height. The rectangle drawn on $[1,2]$ was made using the Midpoint Rule, with a height of $f(1.5)$. That rectangle is labeled "MPR."

These are the three most common rules for determining the heights of approximating rectangles, but one is not forced to use one of these three methods. The rectangle on $[3,4]$ has a height of approximately $f(3.53)$, very close to the Midpoint Rule. It was chosen so that the area of the rectangle is *exactly* the area of the region under f on $[3,4]$. (Later you'll be able to figure how to do this, too.)

The following example will approximate the value of $\int_0^4 (4x-x^2)dx$ using these rules.

Example $5.3.1$: Using the Left Hand, Right Hand and Midpoint Rules

Approximate the value of $\int_0^4 (4x-x^2)dx$ using the Left Hand Rule, the Right Hand Rule, and the Midpoint Rule, using 4 equally spaced subintervals.

Solution

{We break the interval $[0,4]$ into four subintervals as before. In Figure $5.3.3$ we see 4 rectangles drawn on $f(x)=4x-x^2$ using the Left Hand Rule. (The areas of the rectangles are given in each figure.)

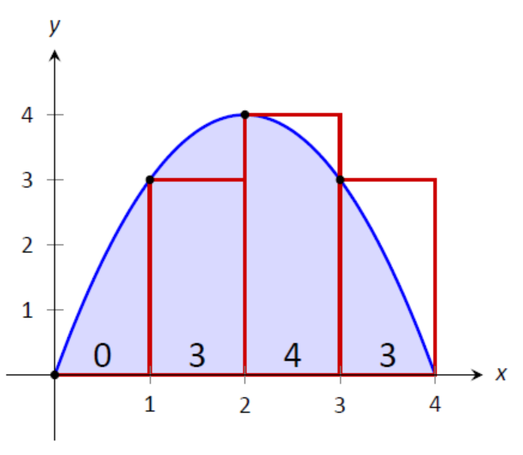


***Figure** $5.3.3$**: Approximating $\int_0^4 (4x-x^2)dx$ using the Left Hand Rule in Example $5.3.1$*

Note how in the first subinterval, $[0,1]$, the rectangle has height $f(0)=0$. We add up the areas of each rectangle (height× width) for our Left Hand Rule approximation:

$$f(0)\cdot 1+f(1)\cdot 1+f(2)\cdot 1+f(3)\cdot 1= \tag{5.3.1}$$
$$0+3+4+3=10. \tag{5.3.2}$$

Figure $5.3.4$ shows 4 rectangles drawn under f using the Right Hand Rule; note how the $[3,4]$ subinterval has a rectangle of height 0.

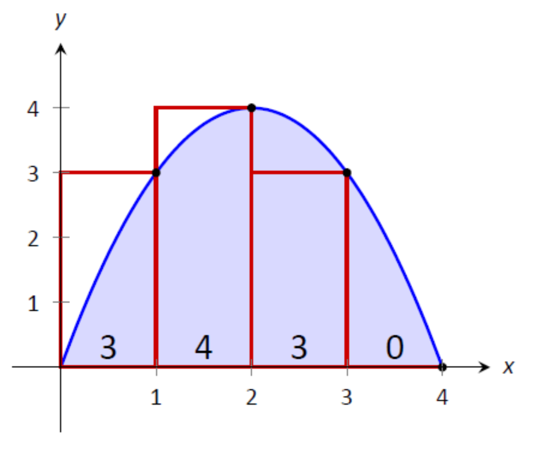


Figure 5.3.4***:*** *Approximating* $\int_0^4 (4x-x^2)dx$ *using the Right Hand Rule in Example* 5.3.1

In this example, these rectangle seem to be the mirror image of those found in Figure 5.3.3. (This is because of the symmetry of our shaded region.) Our approximation gives the same answer as before, though calculated a different way:

$$f(1)\cdot 1+f(2)\cdot 1+f(3)\cdot 1+f(4)\cdot 1= \tag{5.3.3}$$

$$3+4+3+0=10. \tag{5.3.4}$$

Figure 5.3.5 shows 4 rectangles drawn under f using the Midpoint Rule.

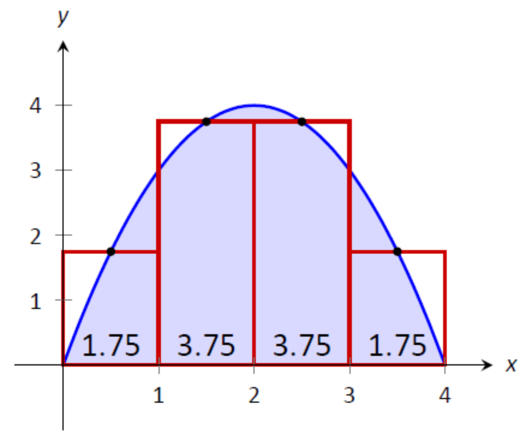


Figure 5.3.5***:*** *Approximating* $\int_0^4 (4x-x^2)dx$ *using the Midpoint Rule in Example* 5.3.1

This gives an approximation of $\int_0^4 (4x-x^2)dx$ as:

$$f(0.5)\cdot 1+f(1.5)\cdot 1+f(2.5)\cdot 1+f(3.5)\cdot 1= \tag{5.3.5}$$

$$1.75+3.75+3.75+1.75=11. \tag{5.3.6}$$

Our three methods provide two approximations of $\int_0^4 (4x-x^2)dx$: 10 and 11.

Summation Notation

It is hard to tell at this moment which is a better approximation: 10 or 11? We can continue to refine our approximation by using more rectangles. The notation can become unwieldy, though, as we add up longer and longer lists of numbers. We introduce **summation notation** to ameliorate this problem.

Suppose we wish to add up a list of numbers a_1, a_2, a_3, \ldots, a_9. Instead of writing

$$a_1+a_2+a_3+a_4+a_5+a_6+a_7+a_8+a_9,\]

we use summation notation and write

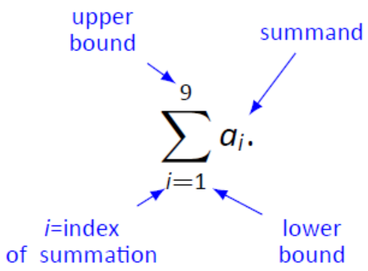


Figure** 5.3.6**: Understanding summation notation

The upper case sigma represents the term "sum." The index of summation in this example is i; any symbol can be used. By convention, the index takes on only the integer values between (and including) the lower and upper bounds.

Let's practice using this notation.

Example 5.3.2: Using summation notation

Let the numbers $\{a_i\}$ be defined as $a_i = 2i - 1$ for integers i, where $i \geq 1$. So $a_1 = 1$, $a_2 = 3$, $a_3 = 5$, etc. (The output is the positive odd integers). Evaluate the following summations:

$$1.\ \sum_{i=1}^{6} a_i \qquad 2.\ \sum_{i=3}^{7} (3a_i - 4) \qquad 3.\ \sum_{i=1}^{4} (a_i)^2 \tag{5.3.7}$$

Solution

1.

$$\sum_{i=1}^{6} a_i = a_1 + a_2 + a_3 + a_4 + a_5 + a_6 \tag{5.3.8}$$
$$= 1 + 3 + 5 + 7 + 9 + 11 \tag{5.3.9}$$
$$= 36. \tag{5.3.10}$$

2. Note the starting value is different than 1:

$$\sum_{i=3}^{7} a_i = (3a_3 - 4) + (3a_4 - 4) + (3a_5 - 4) + (3a_6 - 4) + (3a_7 - 4) \tag{5.3.11}$$
$$= 11 + 17 + 23 + 29 + 35 \tag{5.3.12}$$
$$= 115. \tag{5.3.13}$$

3.

$$\sum_{i=1}^{4} (a_i)^2 = (a_1)^2 + (a_2)^2 + (a_3)^2 + (a_4)^2 \tag{5.3.14}$$
$$= 1^2 + 3^2 + 5^2 + 7^2 \tag{5.3.15}$$
$$= 84 \tag{5.3.16}$$

It might seem odd to stress a new, concise way of writing summations only to write each term out as we add them up. It is. The following theorem gives some of the properties of summations that allow us to work with them without writing individual terms. Examples will follow.

Theorem 5.3.1: Properties of Summations

1. $\sum_{i=1}^{n} c = c \cdot n$, where c is a constant.
2. $\sum_{i=m}^{n} (a_i \pm b_i) = \sum_{i=m}^{n} a_i \pm \sum_{i=m}^{n} b_i$
3. $\sum_{i=m}^{n} c \cdot a_i = c \cdot \sum_{i=m}^{n} a_i$
4. $\sum_{i=m}^{j} a_i + \sum_{i=j+1}^{n} a_i = \sum_{i=m}^{n} a_i$
5. $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$
6. $\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$
7. $\sum_{i=1}^{n} i^3 = \left(\frac{n(n+1)}{2}\right)^2$

Example $5.3.3$: Evaluating summations using Theorem$5.3.1$

Revisit Example $5.3.2$ and, using Theorem $5.3.1$, evaluate

$$\sum_{i=1}^{6} a_i = \sum_{i=1}^{6} (2i-1). \tag{5.3.17}$$

Solution

$$\begin{aligned} \sum_{i=1}^{6}(2i-1) &= \sum_{i=1}^{6} 2i - \sum_{i=1}^{6}(1) && (5.3.18)\\ &= \left(2\sum_{i=1}^{6} i\right) - 6 && (5.3.19)\\ &= 2\frac{6(6+1)}{2} - 6 && (5.3.20)\\ &= 42 - 6 = 36 && (5.3.21) \end{aligned}$$

We obtained the same answer without writing out all six terms. When dealing with small sizes of n, it may be faster to write the terms out by hand. However, Theorem $5.3.1$ is incredibly important when dealing with large sums as we'll soon see.

Riemann Sums

Consider again $\int_0^4 (4x-x^2)dx$. We will approximate this definite integral using 16 equally spaced subintervals and the Right Hand Rule in Example $5.3.4$. Before doing so, it will pay to do some careful preparation.

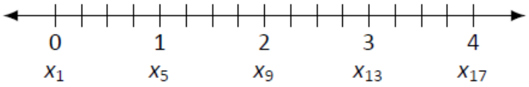
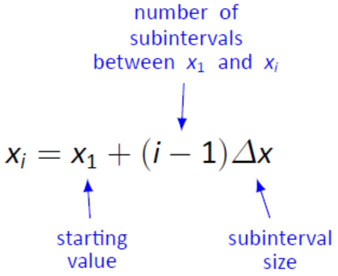


***Figure** $5.3.7$**:** Dividing $[0,4]$ into 16 equally spaced subintervals.*

Figure $5.3.7$ shows a number line of $[0,4]$ divided into 16 equally spaced subintervals. We denote 0 as x_1; we have marked the values of x_5, x_9, x_{13} and x_{17}. We could mark them all, but the figure would get crowded. While it is easy to figure that $x_{10} = 2.25$, in general, we want a method of determining the value of x_i without consulting the figure. Consider:


number of subintervals between x_1 and x_i
$x_i = x_1 + (i-1)\Delta x$
starting value
subinterval size

So $x_{10} = x_1 + 9(4/16) = 2.25$.

If we had partitioned $[0,4]$ into 100 equally spaced subintervals, each subinterval would have length $\Delta x = 4/100 = 0.04$. We could compute x_{32} as

$$x_{32} = x_1 + 31(4/100) = 1.24.\]

(That was far faster than creating a sketch first.)

Given any subdivision of $[0,4]$, the first subinterval is $[x_1, x_2]$; the second is $[x_2, x_3]$; the i th subinterval is $[x_i, x_{i+1}]$.

When using the Left Hand Rule, the height of the i th rectangle will be $f(x_i)$.

When using the Right Hand Rule, the height of the i th rectangle will be $f(x_{i+1})$.

When using the Midpoint Rule, the height of the i th rectangle will be $f\left(\frac{x_i+x_{i+1}}{2}\right)$.

Thus approximating $\int_0^4 (4x-x^2)dx$ with 16 equally spaced subintervals can be expressed as follows, where $\Delta x = 4/16 = 1/4$:

Left Hand Rule: $\sum_{i=1}^{16} f(x_i)\Delta x$

Right Hand Rule: $\sum_{i=1}^{16} f(x_{i+1})\Delta x$

Midpoint Rule: $\sum_{i=1}^{16} f\left(\frac{x_i+x_{i+1}}{2}\right)\Delta x$

We use these formulas in the next two examples. The following example lets us practice using the Right Hand Rule and the summation formulas introduced in Theorem $5.3.1$

Example $5.3.4$: Approximating definite integrals using sums

Approximate $\int_0^4 (4x-x^2)dx$ using the Right Hand Rule and summation formulas with 16 and 1000 equally spaced intervals.

Solution

Using the formula derived before, using 16 equally spaced intervals and the Right Hand Rule, we can approximate the definite integral as

$$\sum_{i=1}^{16} f(x_{i+1})\Delta x. \tag{5.3.22}$$

We have $\Delta x = 4/16 = 0.25$. Since $x_i = 0 + (i-1)\Delta x$, we have

$$\begin{aligned} x_{i+1} &= 0 + \big((i+1)-1\big)\Delta x && (5.3.23)\\ &= i\Delta x && (5.3.24)\end{aligned}$$

Using the summation formulas, consider:

$$\begin{aligned} \int_0^4 (4x-x^2)dx &\approx \sum_{i=1}^{16} f(x_{i+1})\Delta x && (5.3.25)\\ &= \sum_{i=1}^{16} f(i\Delta x)\Delta x && (5.3.26)\\ &= \sum_{i=1}^{16} \big(4i\Delta x-(i\Delta x)^2\big)\Delta x && (5.3.27)\\ &= \sum_{i=1}^{16} (4i\Delta x^2 - i^2\Delta x^3) && (5.3.28)\\ &= (4\Delta x^2)\sum_{i=1}^{16} i - \Delta x^3 \sum_{i=1}^{16} i^2 && (5.3.29)\\ &= (4\Delta x^2)\frac{16\cdot 17}{2} - \Delta x^3 \frac{16(17)(33)}{6} && (5.3.30)\\ &= 4\cdot 0.25^2\cdot 136 - 0.25^3\cdot 1496 && (5.3.31)\\ &= 10.625 && (5.3.32)\end{aligned}$$

We were able to sum up the areas of 16 rectangles with very little computation. In Figure $5.3.8$ the function and the 16 rectangles are graphed. While some rectangles over--approximate the area, other under--approximate the area (by about the same amount). Thus our approximate area of 10.625 is likely a fairly good approximation.

Notice Equation $5.3.31$; by changing the 16's to 1,000's (and appropriately changing the value of Δx), we can use that equation to sum up 1000 rectangles!

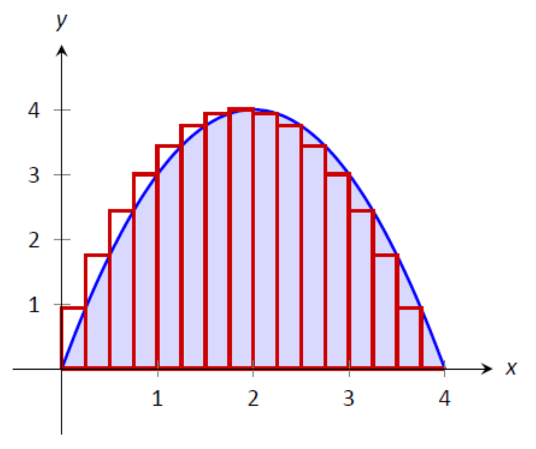


***Figure** 5.3.8: Approximating $\int_0^4(4x-x^2)dx$ with the Right Hand Rule and 16 evenly spaced subintervals.*

We do so here, skipping from the original summand to the equivalent of Equation 5.3.31 to save space. Note that $\Delta x = 4/1000 = 0.004$.

$$\begin{aligned}\int_0^4(4x-x^2)dx &\approx \sum_{i=1}^{1000} f(x_{i+1})\Delta x & (5.3.33)\\ &= (4\Delta x^2)\sum_{i=1}^{1000} i - \Delta x^3 \sum_{i=1}^{1000} i^2 & (5.3.34)\\ &= (4\Delta x^2)\frac{1000\cdot 1001}{2} - \Delta x^3\frac{1000(1001)(2001)}{6} & (5.3.35)\\ &= 4\cdot 0.004^2\cdot 500500 - 0.004^3\cdot 333,833,500 & (5.3.36)\\ &= 10.666656 & (5.3.37)\end{aligned}$$

Using many, many rectangles, we have a likely good approximation of $\int_0^4(4x-x^2)dx$. That is,

$$\int_0^4(4x-x^2)dx \approx 10.666656.\]

Before the above example, we stated what the summations for the Left Hand, Right Hand and Midpoint Rules looked like. Each had the same basic structure, which was:

1. each rectangle has the same width, which we referred to as Δx, and
2. each rectangle's height is determined by evaluating f at a particular point in each subinterval. For instance, the Left Hand Rule states that each rectangle's height is determined by evaluating f at the left hand endpoint of the subinterval the rectangle lives on.

One could partition an interval $[a,b]$ with subintervals that did not have the same size. We refer to the length of the first subinterval as Δx_1, the length of the second subinterval as Δx_2, and so on, giving the length of the $i^{\text{ th}}$ subinterval as Δx_i. Also, one could determine each rectangle's height by evaluating f at \emph{any} point in the $i^{\text{ th}}$ subinterval. We refer to the point picked in the first subinterval as c_1, the point picked in the second subinterval as c_2, and so on, with c_i representing the point picked in the $i^{\text{ th}}$ subinterval. Thus the height of the $i^{\text{ th}}$ subinterval would be $f(c_i)$, and the area of the $i^{\text{ th}}$ rectangle would be $f(c_i)\Delta x_i$.

Summations of rectangles with area $f(c_i)\Delta x_i$ are named after mathematician Georg Friedrich Bernhard Riemann, as given in the following definition.

Definition 5.3.1: Riemann Sum

Let f be defined on the closed interval $[a,b]$ and let Δx be a partition of $[a,b]$, with

$$a=x_1 < x_2 < \ldots < x_n < x_{n+1}=b.\]

Let Δx_i denote the length of the $i^{\text{ th}}$ subinterval $[x_i, x_{i+1}]$ and let c_i denote any value in the $i^{\text{ th}}$ subinterval.

The sum

$$\sum_{i=1}^n f(c_i)\Delta x_i\]

is a **Riemann sum** of f on $[a, b]$.

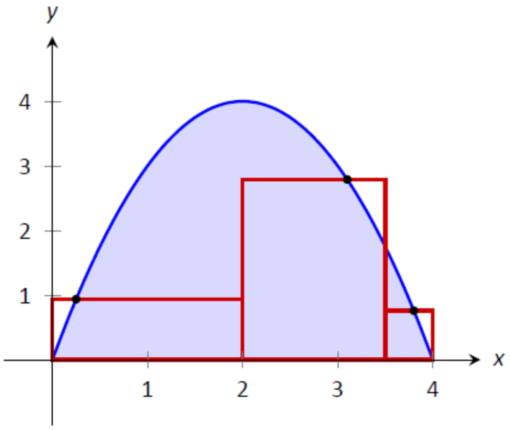


***Figure** $5.3.9$: An example of a general Riemann sum to approximate* $\int_0^4 (4x - x^2)dx$

Figure $5.3.9$ shows the approximating rectangles of a Riemann sum of $\int_0^4 (4x - x^2)dx$. While the rectangles in this example do not approximate well the shaded area, they demonstrate that the subinterval widths may vary and the heights of the rectangles can be determined without following a particular rule.

"Usually" Riemann sums are calculated using one of the three methods we have introduced. The uniformity of construction makes computations easier. Before working another example, let's summarize some of what we have learned in a convenient way.

Key idea 8: Riemann Sum Concepts

Consider $\int_a^b f(x)dx \approx \sum_{i=1}^n f(c_i)\Delta x_i$.

1. When the n subintervals have equal length, $\Delta x_i = \Delta x = \frac{b-a}{n}$.
2. The i $^{\text{th}}$ term of the partition is $x_i = a + (i-1)\Delta x$. (This makes $x_{n+1} = b$.)
3. The Left Hand Rule summation is: $\sum_{i=1}^n f(x_i)\Delta x$.
4. The Right Hand Rule summation is: $\sum_{i=1}^n f(x_{i+1})\Delta x$.
5. The Midpoint Rule summation is: $\sum_{i=1}^n f\left(\frac{x_i + x_{x+1}}{2}\right)\Delta x$.

Let's do another example.

Example $5.3.5$: Approximating definite integrals with sums

Approximate $\int_{-2}^3 (5x+2)dx$ using the Midpoint Rule and 10 equally spaced intervals.

Solution

Following Key Idea 8, we have

$$\Delta x = \frac{3-(-2)}{10} = 1/2 \quad \text{and} \quad x_i = (-2) + (1/2)(i-1) = i/2 - 5/2. \tag{5.3.38}$$

As we are using the Midpoint Rule, we will also need x_{i+1} and $\frac{x_i + x_{i+1}}{2}$. Since $x_i = i/2 - 5/2$, $x_{i+1} = (i+1)/2 - 5/2 = i/2 - 2$. This gives

$$\frac{x_i + x_{i+1}}{2} = \frac{(i/2 - 5/2) + (i/2 - 2)}{2} = \frac{i - 9/2}{2} = i/2 - 9/4. \tag{5.3.39}$$

We now construct the Riemann sum and compute its value using summation formulas.

$$\int_{-2}^3 (5x+2)dx \approx \sum_{i=1}^{10} f\left(\frac{x_i + x_{i+1}}{2}\right)\Delta x \tag{5.3.40}$$

$$= \sum_{i=1}^{10} f(i/2 - 9/4)\Delta x \tag{5.3.41}$$

$$= \sum_{i=1} \big(5(i/2 - 9/4) + 2\big)\Delta x \tag{5.3.42}$$

$$= \Delta x \sum_{i=1}^{10} \left[\left(\frac{5}{2}\right) i - \frac{37}{4} \right] \tag{5.3.43}$$

$$= \Delta x \left(\frac{5}{2} \sum_{i=1}^{10} (i) - \sum_{i=1}^{10} \left(\frac{37}{4}\right) \right) \tag{5.3.44}$$

$$= \frac{1}{2} \left(\frac{5}{2} \cdot \frac{10(11)}{2} - 10 \cdot \frac{37}{4} \right) \tag{5.3.45}$$

$$= \frac{45}{2} = 22.5 \tag{5.3.46}$$

Note the graph of $f(x) = 5x + 2$ in Figure $5.3.10$. The regions whose area is computed by the definite integral are triangles, meaning we can find the exact answer without summation techniques. We find that the exact answer is indeed 22.5. One of the strengths of the Midpoint Rule is that often each rectangle includes area that should not be counted, but misses other area that should. When the partition size is small, these two amounts are about equal and these errors almost "cancel each other out." In this example, since our function is a line, these errors are exactly equal and they do cancel each other out, giving us the exact answer.

Note too that when the function is negative, the rectangles have a "negative" height. When we compute the area of the rectangle, we use $f(c_i)\Delta x$; when f is negative, the area is counted as negative.

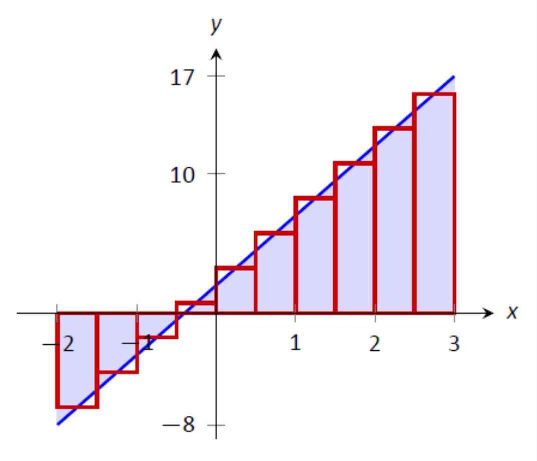


***Figure** $5.3.10$**:** Approximating $\int_{-2}^{3}(5x+2)dx$ using the Midpoint Rule and 10 evenly spaced subintervals in Example $5.3.5$.*

Notice in the previous example that while we used 10 equally spaced intervals, the number "10" didn't play a big role in the calculations until the very end. Mathematicians love to abstract ideas; let's approximate the area of another region using n subintervals, where we do not specify a value of n until the very end.

Example $5.3.6$: Approximating definite integrals with a formula, using sums

Revisit $\int_0^4 (4x - x^2)dx$ yet again. Approximate this definite integral using the Right Hand Rule with n equally spaced subintervals.

Solution

Using Key Idea 8, we know $\Delta x = \frac{4-0}{n} = 4/n$. We also find $x_i = 0 + \Delta x(i-1) = 4(i-1)/n$. The Right Hand Rule uses x_{i+1}, which is $x_{i+1} = 4i/n$.

We construct the Right Hand Rule Riemann sum as follows. Be sure to follow each step carefully. If you get stuck, and do not understand how one line proceeds to the next, you may skip to the result and consider how this result is used. You should come back, though, and work through each step for full understanding.

$$\int_0^4 (4x - x^2)dx \approx \sum_{i=1}^{n} f(x_{i+1})\Delta x \tag{5.3.47}$$

$$= \sum^{n} f\left(\frac{4i}{n}\right)\Delta x \tag{5.3.48}$$

$$= \sum_{i=1}^{n}\left[4\frac{4i}{n}-\left(\frac{4i}{n}\right)^2\right]\Delta x \tag{5.3.49}$$

$$= \sum_{i=1}^{n}\left(\frac{16\Delta x}{n}\right)i-\sum_{i=1}^{n}\left(\frac{16\Delta x}{n^2}\right)i^2 \tag{5.3.50}$$

$$= \left(\frac{16\Delta x}{n}\right)\sum_{i=1}^{n}i-\left(\frac{16\Delta x}{n^2}\right)\sum_{i=1}^{n}i^2 \tag{5.3.51}$$

$$= \left(\frac{16\Delta x}{n}\right)\cdot\frac{n(n+1)}{2} - \left(\frac{16\Delta x}{n^2}\right)\frac{n(n+1)(2n+1)}{6} \qquad \text{(recall } \Delta x = 4/n\text{)} \tag{5.3.52}$$

$$= \frac{32(n+1)}{n}-\frac{32(n+1)(2n+1)}{3n^2} \qquad \text{(now simplify)} \tag{5.3.53}$$

$$= \frac{32}{3}\left(1-\frac{1}{n^2}\right) \tag{5.3.54}$$

The result is an amazing, easy to use formula. To approximate the definite integral with 10 equally spaced subintervals and the Right Hand Rule, set $n = 10$ and compute

$$\int_0^4 (4x-x^2)dx \approx \frac{32}{3}\left(1-\frac{1}{10^2}\right) = 10.56.\]

Recall how earlier we approximated the definite integral with 4 subintervals; with $n = 4$, the formula gives 10, our answer as before.

It is now easy to approximate the integral with 1,000,000 subintervals! Hand-held calculators will round off the answer a bit prematurely giving an answer of 10.66666667. (The actual answer is 10.666666666656.)

We now take an important leap. Up to this point, our mathematics has been limited to geometry and algebra (finding areas and manipulating expressions). Now we apply \textit{calculus}. For any \textit{finite} n, we know that

$$\int_0^4 (4x-x^2)dx \approx \frac{32}{3}\left(1-\frac{1}{n^2}\right).\]

Both common sense and high--level mathematics tell us that as n gets large, the approximation gets better. In fact, if we take the *limit* as $n \to \infty$, we get the *exact area* described by $\int_0^4 (4x-x^2)dx$. That is,

$$\int_0^4 (4x-x^2)dx = \lim_{n\to\infty}\frac{32}{3}\left(1-\frac{1}{n^2}\right) \tag{5.3.55}$$

$$= \frac{32}{3}(1-0) \tag{5.3.56}$$

$$= \frac{32}{3} = 10.\overline{6} \tag{5.3.57}$$

This is a fantastic result. By considering n equally--spaced subintervals, we obtained a formula for an approximation of the definite integral that involved our variable n. As n grows large -- without bound -- the error shrinks to zero and we obtain the exact area.

This section started with a fundamental calculus technique: make an approximation, refine the approximation to make it better, then use limits in the refining process to get an exact answer. That is precisely what we just did.

Let's practice this again.

Example $5.3.7$: Approximating definite integrals with a formula, using sums

Find a formula that approximates $\int_{-1}^{5} x^3 dx$ using the Right Hand Rule and n equally spaced subintervals, then take the limit as $n \to \infty$ to find the exact area.

Solution

Following Key Idea 8, we have $\Delta x = \frac{5-(-1)}{n} = 6/n$. We have $x_i = (-1) + (i-1)\Delta x$; as the Right Hand Rule uses x_{i+1}, we have $x_{i+1} = (-1) + i\Delta x$.

The Riemann sum corresponding to the Right Hand Rule is (followed by simplifications):

$$\int^{5} x^3 dx \approx \sum^{n} f(x_{i+1})\Delta x \tag{5.3.58}$$

$$= \sum_{i=1}^{n} f(-1+i\Delta x)\Delta x \tag{5.3.59}$$

$$= \sum_{i=1}^{n} (-1+i\Delta x)^3\Delta x \tag{5.3.60}$$

$$= \sum_{i=1}^{n} \Big((i\Delta x)^3 - 3(i\Delta x)^2 + 3i\Delta x - 1\Big)\Delta x \tag{5.3.61}$$

$$\text{(now distribute } \Delta x\text{)}$$

$$= \sum_{i=1}^{n} \left(i^3\Delta x^4 - 3i^2\Delta x^3 + 3i\Delta x^2 - \Delta x\right) \tag{5.3.62}$$

$$\text{(now split up summation)}$$

$$= \Delta x^4 \sum_{i=1}^{n} i^3 - 3\Delta x^3 \sum_{i=1}^{n} i^2 + 3\Delta x^2 \sum_{i=1}^{n} i - \sum_{i=1}^{n} \Delta x \tag{5.3.63}$$

$$= \Delta x^4 \left(\frac{n(n+1)}{2}\right)^2 - 3\Delta x^3 \frac{n(n+1)(2n+1)}{6} + 3\Delta x^2 \frac{n(n+1)}{2} - n\Delta x \tag{5.3.64}$$

$$\text{(use } \Delta x = 6/n\text{)} \tag{5.3.65}$$

$$= \frac{1296}{n^4} \cdot \frac{n^2(n+1)^2}{4} - 3\frac{216}{n^3} \cdot \frac{n(n+1)(2n+1)}{6} + 3\frac{36}{n^2}\frac{n(n+1)}{2} - 6 \tag{5.3.66}$$

$$\text{(now do a sizable amount of algebra to simplify)} \tag{5.3.67}$$

$$= 156 + \frac{378}{n} + \frac{216}{n^2} \tag{5.3.68}$$

Once again, we have found a compact formula for approximating the definite integral with n equally spaced subintervals and the Right Hand Rule. Using 10 subintervals, we have an approximation of 195.96 (these rectangles are shown in Figure $5.3.11$. Using $n = 100$ gives an approximation of 159.802.

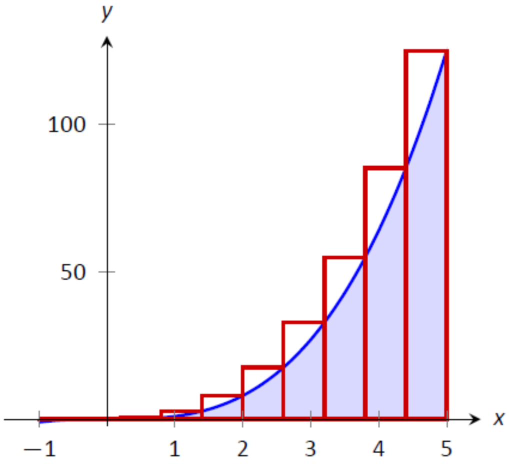


Figure $5.3.11$***:*** *Approximating $\int_{-1}^{5} x^3 dx$ using the Right Hand Rule and 10 evenly spaced subintervals.*

Now find the exact answer using a limit:

$$\int_{-1}^5 x^3dx = \lim_{n\to\infty} \left(156 + \frac{378}n + \frac{216}{n^2}\right) = 156.\]

Limits of Riemann Sums

We have used limits to evaluate exactly given definite limits. Will this always work? We will show, given not--very--restrictive conditions, that yes, it will always work.

The previous two examples demonstrated how an expression such as

$$\sum_{i=1}^n f(x_{i+1})\Delta x\]

can be rewritten as an expression explicitly involving n, such as $32/3(1-1/n^2)$.

Viewed in this manner, we can think of the summation as a function of n. An n value is given (where n is a positive integer), and the sum of areas of n equally spaced rectangles is returned, using the Left Hand, Right Hand, or Midpoint Rules.

Given a definite integral $\int_a^b f(x)dx$, let:

1. $S_L(n)=\sum_{i=1}^n f(x_i)\Delta x$, the sum of equally spaced rectangles formed using the Left Hand Rule,
2. $S_R(n)=\sum_{i=1}^n f(x_{i+1})\Delta x$, the sum of equally spaced rectangles formed using the Right Hand Rule, and
3. $S_M(n)=\sum_{i=1}^n f\left(\frac{x_i+x_{i+1}}{2}\right)\Delta x$, the sum of equally spaced rectangles formed using the Midpoint Rule.

Recall the definition of a limit as $n\to\infty$: $\lim_{n\to\infty} S_L(n)=K$ if, given any $\epsilon>0$, there exists $N>0$ such that

$$\left|S_L(n)-K\right| < \epsilon \quad \text{when}\quad n\geq N.\]

The following theorem states that we can use any of our three rules to find the exact value of a definite integral $\int_a^b f(x)dx$. It also goes two steps further. The theorem states that the height of each rectangle doesn't have to be determined following a specific rule, but could be $f(c_i)$, where c_i is any point in the i $^{\text{th}}$ subinterval, as discussed before Riemann Sums where defined in Definition $5.3.1$.

The theorem goes on to state that the rectangles do not need to be of the same width. Using the notation of Definition $5.3.1$, let Δx_i denote the length of the i $^{\text{th}}$ subinterval in a partition of $[a,b]$. Now let $||\Delta x||$ represent the length of the largest subinterval in the partition: that is, $||\Delta x||$ is the largest of all the Δx_i's. If $||\Delta x||$ is small, then $[a,b]$ must be partitioned into many subintervals, since all subintervals must have small lengths. "Taking the limit as $||\Delta x||$ goes to zero" implies that the number n of subintervals in the partition is growing to infinity, as the largest subinterval length is becoming arbitrarily small. We then interpret the expression

$$\lim_{||\Delta x||\to 0}\sum_{i=1}^nf(c_i)\Delta x_i\]

as "the limit of the sum of rectangles, where the width of each rectangle can be different but getting small, and the height of each rectangle is not necessarily determined by a particular rule." The theorem states that this Riemann Sum also gives the value of the definite integral of f over $[a,b]$.

Theorem $5.3.2$: Definite Integrals and the Limit of Riemann Sums

Let f be continuous on the closed interval $[a,b]$ and let $S_L(n)$, $S_R(n)$ and $S_M(n)$ be defined as before. Then:

1. $\lim_{n\to\infty} S_L(n)=\lim_{n\to\infty} S_R(n)=\lim_{n\to\infty} S_M(n)=\lim_{n\to\infty}\sum_{i=1}^n f(c_i)\Delta x,$
2. $\lim_{n\to\infty}\sum_{i=1}^n f(c_i)\Delta x=\int_a^b f(x)dx\$, and.$
3. $\lim_{\|\Delta x\|\to 0}\sum_{i=1}^n f(c_i)\Delta x_i=\int_a^b f(x)dx.$

We summarize what we have learned over the past few sections here.

- Knowing the "area under the curve" can be useful. One common example is: the area under a velocity curve is displacement.
- We have defined the definite integral, $\int_a^b f(x)dx$, to be the signed area under f on the interval $[a,b]$.
- While we can approximate a definite integral many ways, we have focused on using rectangles whose heights can be determined using: the Left Hand Rule, the Right Hand Rule and the Midpoint Rule.
- Sums of rectangles of this type are called Riemann sums.
- The exact value of the definite integral can be computed using the limit of a Riemann sum. We generally use one of the above methods as it makes the algebra simpler.

We first learned of derivatives through limits then learned rules that made the process simpler. We know of a way to evaluate a definite integral using limits; in the next section we will see how the Fundamental Theorem of Calculus makes the process simpler. The key feature of this theorem is its connection between the indefinite integral and the definite integral.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/
- Integrated by Justin Marshall.

This page titled 5.3: Riemann Sums is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

5.4: The Fundamental Theorem of Calculus

The Fundamental Theorem of Calculus

Let $f(t)$ be a continuous function defined on $[a,b]$. The definite integral $\int_a^b f(x)\,dx$ is the "area under f" on $[a,b]$. We can turn this concept into a function by letting the upper (or lower) bound vary.

Let $F(x)=\int_a^x f(t)\,dt$. It computes the area under f on $[a,x]$ as illustrated in Figure $5.4.1$. We can study this function using our knowledge of the definite integral. For instance, $F(a)=0$ since $\int_a^a f(t)\,dt=0$.

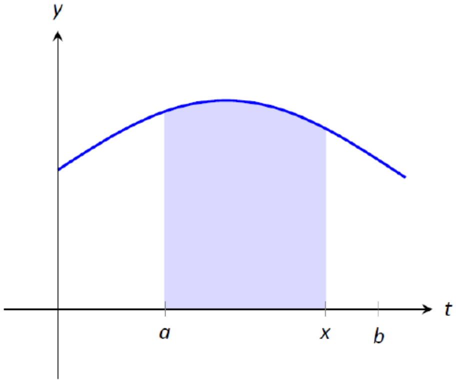


Figure $5.4.1$**:** The area of the shaded region is $F(x)=\int_a^x f(t)\,dt$.

We can also apply calculus ideas to $F(x)$; in particular, we can compute its derivative. While this may seem like an innocuous thing to do, it has far--reaching implications, as demonstrated by the fact that the result is given as an important theorem.

Theorem $5.4.1$: The Fundamental Theorem of Calculus, Part 1

Let f be continuous on $[a,b]$ and let $F(x)=\int_a^x f(t)\,dt$. Then F is a differentiable function on (a,b), and

$$F'(x)=f(x). \tag{5.4.1}$$

Initially this seems simple, as demonstrated in the following example.

Example $5.4.1$: Using the Fundamental Theorem of Calculus, Part 1

Let $F(x)=\int_{-5}^x (t^2+\sin t)\,dt$. What is $F'(x)$?}

Solution

Using the Fundamental Theorem of Calculus, we have $F'(x)=x^2+\sin x$.

This simple example reveals something incredible: $F(x)$ is an antiderivative of $x^2+\sin x$! Therefore, $F(x)=\frac{1}{3}x^3-\cos x+C$ for some value of C. (We can find C, but generally we do not care. We know that $F(-5)=0$, which allows us to compute C. In this case, $C=\cos(-5)+\frac{125}{3}$.)

We have done more than found a complicated way of computing an antiderivative. Consider a function f defined on an open interval containing a, b and c. Suppose we want to compute $\int_a^b f(t)\,dt$. First, let $F(x)=\int_c^x f(t)\,dt$. Using the properties of the definite integral found in Theorem 5.2.1, we know

$$\int_a^b f(t)\,dt=\int_a^c f(t)\,dt+\int_c^b f(t)\,dt \tag{5.4.2}$$

$$=-\int_c^a f(t)\,dt+\int_c^b f(t)\,dt \tag{5.4.3}$$

$$\begin{aligned} &= -F(a) + F(b) && (5.4.4)\\ &= F(b) - F(a). && (5.4.5)\end{aligned}$$

We now see how indefinite integrals and definite integrals are related: we can evaluate a definite integral using antiderivatives! This is the second part of the Fundamental Theorem of Calculus.

Theorem $5.4.2$: The Fundamental Theorem of Calculus, Part 2

Let f be continuous on $[a, b]$ and let F be *any* antiderivative of f. Then

$$\int_a^b f(x)\,dx = F(b) - F(a). \tag{5.4.6}$$

Example $5.4.2$: Using the Fundamental Theorem of Calculus, Part 2

We spent a great deal of time in the previous section studying $\int_0^4 (4x - x^2)\,dx$. Using the Fundamental Theorem of Calculus, evaluate this definite integral.

Solution

We need an antiderivative of $f(x) = 4x - x^2$. All antiderivatives of f have the form $F(x) = 2x^2 - \frac{1}{3}x^3 + C$; for simplicity, choose $C = 0$.

The Fundamental Theorem of Calculus states

$$\int_0^4 (4x - x^2)\,dx = F(4) - F(0) = \big(2(4)^2 - \frac{1}{3}4^3\big) - \big(0 - 0\big) = 32 - \frac{64}{3} = 32/3. \tag{5.4.7}$$

This is the same answer we obtained using limits in the previous section, just with much less work.

Notation: A special notation is often used in the process of evaluating definite integrals using the Fundamental Theorem of Calculus. Instead of explicitly writing $F(b) - F(a)$, the notation $F(x)\Big|_a^b$ is used. Thus the solution to Example $5.4.2$ would be written as:

$$\int_0^4 (4x - x^2)\,dx = \left(2x^2 - \frac{1}{3}x^3\right)\Big|_0^4 = \big(2(4)^2 - \frac{1}{3}4^3\big) - \big(0 - 0\big) = 32/3. \tag{5.4.8}$$

The Constant C: *Any* antiderivative $F(x)$ can be chosen when using the Fundamental Theorem of Calculus to evaluate a definite integral, meaning any value of C can be picked. The constant *always* cancels out of the expression when evaluating $F(b) - F(a)$, so it does not matter what value is picked. This being the case, we might as well let $C = 0$.

Example $5.4.3$: Using the Fundamental Theorem of Calculus, Part 2

Evaluate the following definite integrals.

$$1.\ \int_{-2}^2 x^3\,dx \quad 2.\ \int_0^\pi \sin x\,dx \quad 3.\ \int_0^5 e^t\,dt \quad 4.\ \int_4^9 \sqrt{u}\,du \quad 5.\ \int_1^5 2\,dx \tag{5.4.9}$$

Solution

1. $\displaystyle\int_{-2}^2 x^3\,dx = \frac{1}{4}x^4\Big|_{-2}^2 = \left(\frac{1}{4}2^4\right) - \left(\frac{1}{4}(-2)^4\right) = 0.$
2. $\displaystyle\int_0^\pi \sin x\,dx = -\cos x\Big|_0^\pi = -\cos\pi - \big(-\cos 0\big) = 1 + 1 = 2.$
   (This is interesting; it says that the area under one "hump" of a sine curve is 2.)
3. $\displaystyle\int_0^5 e^t\,dt = e^t\Big|_0^5 = e^5 - e^0 = e^5 - 1 \approx 147.41.$
4. $\displaystyle\int_4^9 \sqrt{u}\,du = \int_4^9 u^{\frac{1}{2}}\,du = \frac{2}{3}u^{\frac{3}{2}}\Big|_4^9 = \frac{2}{3}\left(9^{\frac{3}{2}} - 4^{\frac{3}{2}}\right) = \frac{2}{3}\big(27 - 8\big) = \frac{38}{3}.$
5. $\displaystyle\int_1^5 2\,dx = 2x\Big|_1^5 = 2(5) - 2 = 2(5 - 1) = 8.$

This integral is interesting; the integrand is a constant function, hence we are finding the area of a rectangle with width $(5 - 1) = 4$ and height 2. Notice how the evaluation of the definite integral led to $2(4) = 8$.

In general, if c is a constant, then $\int_a^b c\,dx = c(b-a)$.

Understanding Motion with the Fundamental Theorem of Calculus

We established, starting with Key Idea 1, that the derivative of a position function is a velocity function, and the derivative of a velocity function is an acceleration function. Now consider definite integrals of velocity and acceleration functions. Specifically, if $v(t)$ is a velocity function, what does $\int_a^b v(t)\,dt$ mean?

The Fundamental Theorem of Calculus states that

$$\int_a^b v(t)\,dt = V(b) - V(a), \tag{5.4.10}$$

where $V(t)$ is any antiderivative of $v(t)$. Since $v(t)$ is a velocity function, $V(t)$ must be a position function, and $V(b) - V(a)$ measures a change in position, or **displacement**.

Example $5.4.4$: Finding displacement

A ball is thrown straight up with velocity given by $v(t) = -32t + 20\text{ft/s}$, where t is measured in seconds. Find, and interpret, $\int_0^1 v(t)\,dt.\}$

Solution

Using the Fundamental Theorem of Calculus, we have

$$\begin{aligned} \int_0^1 v(t)\,dt &= \int_0^1 (-32t+20)\,dt \qquad (5.4.11)\\ &= -16t^2 + 20t\Big|_0^1 \qquad (5.4.12)\\ &= 4. \qquad (5.4.13)\end{aligned}$$

Thus if a ball is thrown straight up into the air with velocity $v(t) = -32t + 20$, the height of the ball, 1 second later, will be 4 feet above the initial height. (Note that the ball has *traveled* much farther. It has gone up to its peak and is falling down, but the difference between its height at $t = 0$ and $t = 1$ is 4 ft.

Integrating a rate of change function gives total change. Velocity is the rate of position change; integrating velocity gives the total change of position, i.e., displacement.

Integrating a speed function gives a similar, though different, result. Speed is also the rate of position change, but does not account for direction. So integrating a speed function gives total change of position, without the possibility of "negative position change." Hence the integral of a speed function gives *distance traveled.*

As acceleration is the rate of velocity change, integrating an acceleration function gives total change in velocity. We do not have a simple term for this analogous to displacement. If $a(t) = 5\text{ miles/h}^2$ and t is measured in hours, then

$$\int_0^3 a(t)\,dt = 15 \tag{5.4.14}$$

means the velocity has increased by 15 m/h from $t = 0$ to $t = 3$.

The Fundamental Theorem of Calculus and the Chain Rule

Part 1 of the Fundamental Theorem of Calculus (FTC) states that given $F(x) = \int_a^x f(t)\,dt$, $F'(x) = f(x)$. Using other notation, $\frac{d}{dx}\big(F(x)\big) = f(x)$. While we have just practiced evaluating definite integrals, sometimes finding antiderivatives is impossible and we need to rely on other techniques to approximate the value of a definite integral. Functions written as $F(x) = \int_a^x f(t)\,dt$ are useful in such situations.

It may be of further use to compose such a function with another. As an example, we may compose $F(x)$ with $g(x)$ to get

$$F(g(x)) = \int^{g(x)} f(t)\,dt \tag{5.4.15}$$

$$F(g(x)) = \int_a f(t)\,dt. \tag{5.4.15}$$

What is the derivative of such a function? The Chain Rule can be employed to state

$$\frac{d}{dx}\Big(F\big(g(x)\big)\Big) = F'\big(g(x)\big)g'(x) = f\big(g(x)\big)g'(x). \tag{5.4.16}$$

An example will help us understand this.

Example 5.4.4: The FTC, Part 1, and the Chain Rule

Find the derivative of $F(x) = \int_2^{x^2} \ln t\,dt$.

Solution

We can view $F(x)$ as being the function $G(x) = \int_2^x \ln t\,dt$ composed with $g(x) = x^2$; that is, $F(x) = G\big(g(x)\big)$. The Fundamental Theorem of Calculus states that $G'(x) = \ln x$. The Chain Rule gives us

$$\begin{aligned} F'(x) &= G'\big(g(x)\big)g'(x) && (5.4.17)\\ &= \ln(g(x))g'(x) && (5.4.18)\\ &= \ln(x^2)2x && (5.4.19)\\ &= 2x\ln x^2 && (5.4.20) \end{aligned}$$

Normally, the steps defining $G(x)$ and $g(x)$ are skipped.

Practice this once more.

Example 5.4.5: The FTC, Part 1, and the Chain Rule

Find the derivative of $F(x) = \int_{\cos x}^{5} t^3\,dt$.

Solution

Note that $F(x) = -\int_5^{\cos x} t^3\,dt$. Viewed this way, the derivative of F is straightforward:

$$F'(x) = \sin x\cos^3 x. \tag{5.4.21}$$

Area Between Curves

Consider continuous functions $f(x)$ and $g(x)$ defined on $[a,b]$, where $f(x) \ge g(x)$ for all x in $[a,b]$, as demonstrated in Figure $5.4.2$. What is the area of the shaded region bounded by the two curves over $[a,b]$?

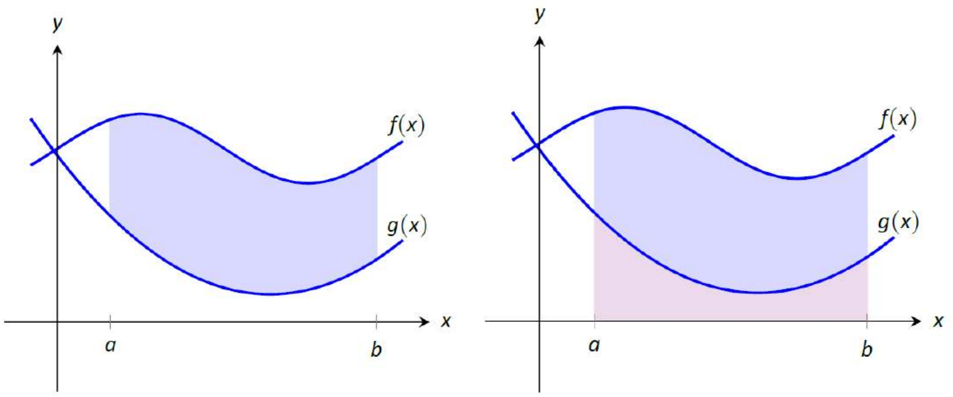


Figure $5.4.2$**:** Finding the area bounded by two functions on an interval; it is found by subtracting the area under g from the area under f.

The area can be found by recognizing that this area is "the area under f $-$ the area under g." Using mathematical notation, the area is

$$\int_a^b f(x)\,dx - \int_a^b g(x)\,dx. \tag{5.4.22}$$

Properties of the definite integral allow us to simplify this expression to

$$\int_a^b \big(f(x) - g(x)\big)\,dx. \tag{5.4.23}$$

Theorem 5.4.3: Area Between Curves

Let $f(x)$ and $g(x)$ be continuous functions defined on $[a,b]$ where $f(x) \geq g(x)$ for all x in $[a,b]$. The area of the region bounded by the curves $y = f(x)$, $y = g(x)$ and the lines $x = a$ and $x = b$ is

$$\int_a^b \big(f(x) - g(x)\big)\,dx. \tag{5.4.24}$$

Example 5.4.6: Finding area between curves

Find the area of the region enclosed by $y = x^2 + x - 5$ and $y = 3x - 2$.

Solution

It will help to sketch these two functions, as done in Figure 5.4.3.

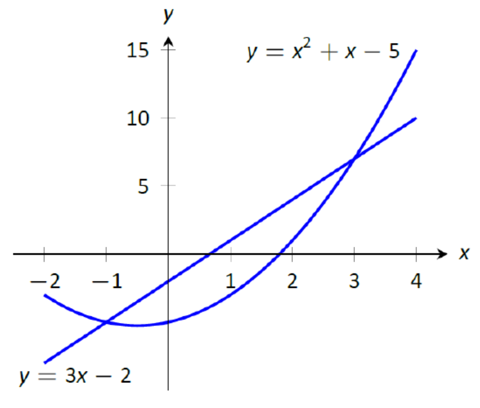


Figure 5.4.3: Sketching the region enclosed by $y = x^2 + x - 5$ and $y = 3x - 2$ in Example 5.4.6

The region whose area we seek is completely bounded by these two functions; they seem to intersect at $x = -1$ and $x = 3$. To check, set $x^2 + x - 5 = 3x - 2$ and solve for x:

$$x^2 + x - 5 = 3x - 2 \tag{5.4.25}$$

$$(x^2 + x - 5) - (3x - 2) = 0 \tag{5.4.26}$$

$$x^2 - 2x - 3 = 0 \tag{5.4.27}$$

$$(x-3)(x+1) = 0 \tag{5.4.28}$$

$$x = -1,\ 3. \tag{5.4.29}$$

Following Theorem 5.4.3, the area is

$$\int_{-1}^{3} \big(3x - 2 - (x^2 + x - 5)\big)\,dx = \int_{-1}^{3} (-x^2 + 2x + 3)\,dx \tag{5.4.30}$$

$$= \left(-\frac{1}{3}x^3 + x^2 + 3x\right)\Big|_{-1}^{3} \tag{5.4.31}$$

$$= -\frac{1}{3}(27) + 9 + 9 - \left(\frac{1}{3} + 1 - 3\right) \tag{5.4.32}$$

$$= 10\frac{2}{3} = 10.\overline{6} \tag{5.4.33}$$

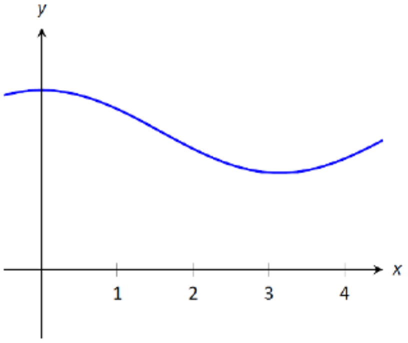


Figure $5.4.4$**:** A graph of a function f to introduce the Mean Value Theorem.

Consider the graph of a function f in Figure $5.4.4$ and the area defined by $\int_1^4 f(x)\,dx$. Three rectangles are drawn in Figure $5.4.5$; in (a), the height of the rectangle is greater than f on $[1,4]$, hence the area of this rectangle is is greater than $\int_0^4 f(x)\,dx$.

In (b), the height of the rectangle is smaller than f on $[1,4]$, hence the area of this rectangle is less than $\int_1^4 f(x)\,dx$.

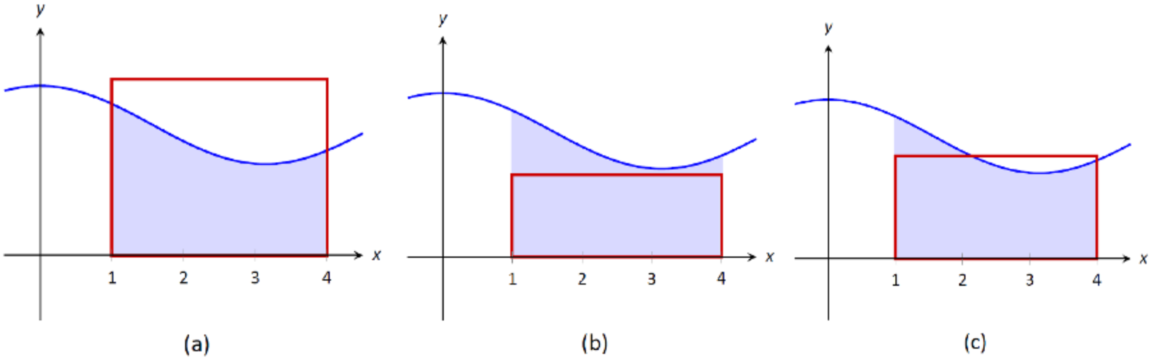


Figure $5.4.5$**:** Differently sized rectangles give upper and lower bounds on $\int_1^4 f(x)\,dx$; the last rectangle matches the area exactly.

Finally, in (c) the height of the rectangle is such that the area of the rectangle is *exactly* that of $\int_0^4 f(x)\,dx$. Since rectangles that are "too big", as in (a), and rectangles that are "too little," as in (b), give areas greater/lesser than $\int_1^4 f(x)\,dx$, it makes sense that there is a rectangle, whose top intersects $f(x)$ somewhere on $[1,4]$, whose area is *exactly* that of the definite integral.

We state this idea formally in a theorem.

Theorem $5.4.4$: The Mean Value Theorem of Integration

Let f be continuous on $[a,b]$. There exists a value c in $[a,b]$ such that

$$\int_a^b f(x)\,dx = f(c)(b-a). \tag{5.4.34}$$

This is an *existential* statement; c exists, but we do not provide a method of finding it. Theorem $5.4.4$ is directly connected to the Mean Value Theorem of Differentiation, given as Theorem 3.2.1; we leave it to the reader to see how.

We demonstrate the principles involved in this version of the Mean Value Theorem in the following example.

Example 5.4.7: Using the Mean Value Theorem

Consider $\int_0^\pi \sin x\,dx$. Find a value c guaranteed by the Mean Value Theorem.

Solution

We first need to evaluate $\int_0^\pi \sin x\,dx$. (This was previously done in Example 5.4.3)

$$\int_0^\pi \sin x\,dx = -\cos x\Big|_0^\pi = 2. \tag{5.4.35}$$

Thus we seek a value c in $[0,\pi]$ such that $\pi\sin c = 2$.

$$\pi\sin c = 2 \quad\Rightarrow\quad \sin c = 2/\pi \quad\Rightarrow\quad c = \arcsin(2/\pi) \approx 0.69. \tag{5.4.36}$$

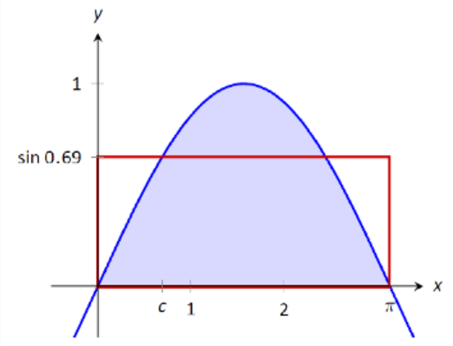


Figure 5.4.6**:** A graph of $y = \sin x$ on $[0,\pi]$ and the rectangle guaranteed by the Mean Value Theorem.

In Figure 5.4.6 $\sin x$ is sketched along with a rectangle with height $\sin(0.69)$. The area of the rectangle is the same as the area under $\sin x$ on $[0,\pi]$.

Let f be a function on $[a,b]$ with c such that $f(c)(b-a) = \int_a^b f(x)\,dx$. Consider $\int_a^b \big(f(x)-f(c)\big)\,dx$:

$$\begin{aligned}\int_a^b \big(f(x)-f(c)\big)\,dx &= \int_a^b f(x) - \int_a^b f(c)\,dx &(5.4.37)\\ &= f(c)(b-a) - f(c)(b-a) &(5.4.38)\\ &= 0. &(5.4.39)\end{aligned}$$

When $f(x)$ is shifted by $-f(c)$, the amount of area under f above the x-axis on $[a,b]$ is the same as the amount of area below the x-axis above f; see Figure 5.4.7 for an illustration of this. In this sense, we can say that $f(c)$ is the *average value* of f on $[a,b]$.

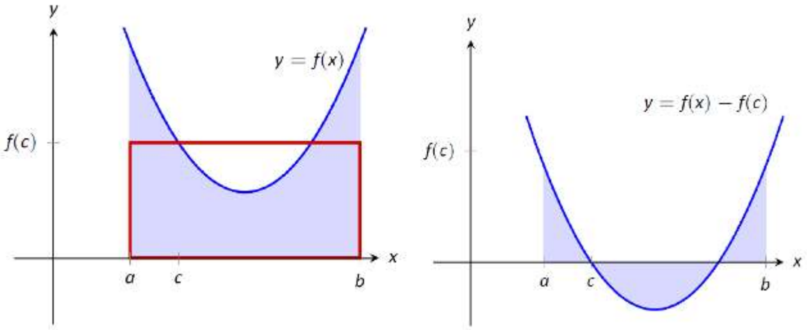


Figure 5.4.7**:** On the left, a graph of $y = f(x)$ and the rectangle guaranteed by the Mean Value Theorem. On the right, $y = f(x)$ is shifted down by $f(c)$; the resulting "area under the curve" is 0.

The value $f(c)$ is the average value in another sense. First, recognize that the Mean Value Theorem can be rewritten as

$$f(c) = \frac{1}{b-a}\int_a^b f(x)\,dx, \tag{5.4.40}$$

for some value of c in $[a, b]$. Next, partition the interval $[a, b]$ into n equally spaced subintervals, $a = x_1 < x_2 < \ldots < x_{n+1} = b$ and choose any c_i in $[x_i, x_{i+1}]$. The average of the numbers $f(c_1)$, $f(c_2)$, ..., $f(c_n)$ is:

$$\frac{1}{n}\Big(f(c_1) + f(c_2) + \ldots + f(c_n)\Big) = \frac{1}{n}\sum_{i=1}^{n} f(c_i). \tag{5.4.41}$$

Multiply this last expression by 1 in the form of $\frac{(b-a)}{(b-a)}$:

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} f(c_i) &= \sum_{i=1}^{n} f(c_i)\frac{1}{n} && (5.4.42) \\ &= \sum_{i=1}^{n} f(c_i)\frac{1}{n}\frac{(b-a)}{(b-a)} && (5.4.43) \\ &= \frac{1}{b-a}\sum_{i=1}^{n} f(c_i)\,\Delta x \quad (\text{where } \Delta x = (b-a)/n) && (5.4.44) \end{aligned}$$

Now take the limit as $n \to \infty$:

$$\lim_{n\to\infty} \frac{1}{b-a}\sum_{i=1}^{n} f(c_i)\,\Delta x \quad = \quad \frac{1}{b-a}\int_a^b f(x)\,dx \quad = \quad f(c). \tag{5.4.45}$$

This tells us this: when we evaluate f at n (somewhat) equally spaced points in $[a, b]$, the average value of these samples is $f(c)$ as $n \to \infty$.

This leads us to a definition.

Definition 5.4.1: The Average Value of f on $[a, b]$

Let f be continuous on $[a, b]$. The **average value of f on** $[a, b]$ is $f(c)$, where c is a value in $[a, b]$ guaranteed by the Mean Value Theorem. I.e.,

$$\text{Average Value of } f \text{ on } [a, b] = \frac{1}{b-a}\int_a^b f(x)\,dx. \tag{5.4.46}$$

An application of this definition is given in the following example.

Example 5.4.8: Finding the average value of a function

An object moves back and forth along a straight line with a velocity given by $v(t) = (t-1)^2$ on $[0, 3]$, where t is measured in seconds and $v(t)$ is measured in ft/s.

What is the average velocity of the object?

Solution

By our definition, the average velocity is:

$$\frac{1}{3-0}\int_0^3 (t-1)^2\,dt = \frac{1}{3}\int_0^3 \Big(t^2 - 2t + 1\Big)\,dt = \frac{1}{3}\left(\frac{1}{3}t^3 - t^2 + t\right)\Big|_0^3 = 1 \text{ ft/s}. \tag{5.4.47}$$

We can understand the above example through a simpler situation. Suppose you drove 100 miles in 2 hours. What was your average speed? The answer is simple: $\text{displacement/time} = 100 \text{ miles}/2 \text{ hours} = 50 mph$.

What was the displacement of the object in Example 5.4.8? We calculate this by integrating its velocity function: $\int_0^3 (t-1)^2\,dt = 3$ ft. Its final position was 3 feet from its initial position after 3 seconds: its average velocity was 1 ft/s.

This section has laid the groundwork for a lot of great mathematics to follow. The most important lesson is this: definite integrals can be evaluated using antiderivatives. Since the previous section established that definite integrals are the limit of Riemann sums, we can later create Riemann sums to approximate values other than "area under the curve," convert the sums to definite integrals, then evaluate these using the Fundamental Theorem of Calculus. This will allow us to compute the work done by a variable force, the volume of certain solids, the arc length of curves, and more.

The downside is this: generally speaking, computing antiderivatives is much more difficult than computing derivatives. The next chapter is devoted to techniques of finding antiderivatives so that a wide variety of definite integrals can be evaluated. Before that, the next section explores techniques of approximating the value of definite integrals beyond using the Left Hand, Right Hand and Midpoint Rules.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/
- Integrated by Justin Marshall.

This page titled 5.4: The Fundamental Theorem of Calculus is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

5.5: Numerical Integration

The Fundamental Theorem of Calculus gives a concrete technique for finding the exact value of a definite integral. That technique is based on computing antiderivatives. Despite the power of this theorem, there are still situations where we must *approximate* the value of the definite integral instead of finding its exact value. The first situation we explore is where we *cannot* compute the antiderivative of the integrand. The second case is when we actually do not know the integrand, but only its value when evaluated at certain points.

An **elementary function** is any function that is a combination of polynomials, n^{th} roots, rational, exponential, logarithmic and trigonometric functions. We can compute the derivative of any elementary function, but there are many elementary functions of which we cannot compute an antiderivative. For example, the following functions do not have antiderivatives that we can express with elementary functions:

$$e^{-x^2}, \quad \sin(x^3)\quad \text{and} \quad \frac{\sin x}{x}.$$

The simplest way to refer to the antiderivatives of e^{-x^2} is to simply write $\int e^{-x^2}\,dx$.

This section outlines three common methods of approximating the value of definite integrals. We describe each as a systematic method of approximating area under a curve. By approximating this area accurately, we find an accurate approximation of the corresponding definite integral.

We will apply the methods we learn in this section to the following definite integrals:

$$\int_0^1 e^{-x^2} \ dx, \quad \int_{-\frac{\pi}{4}}^{\frac{\pi}{2}} \sin(x^3) \ dx, \quad \text{and} \quad \int_{0.5}^{4\pi} \frac{\sin(x)}{x} \ dx,$$

as pictured in Figure 5.5.1.

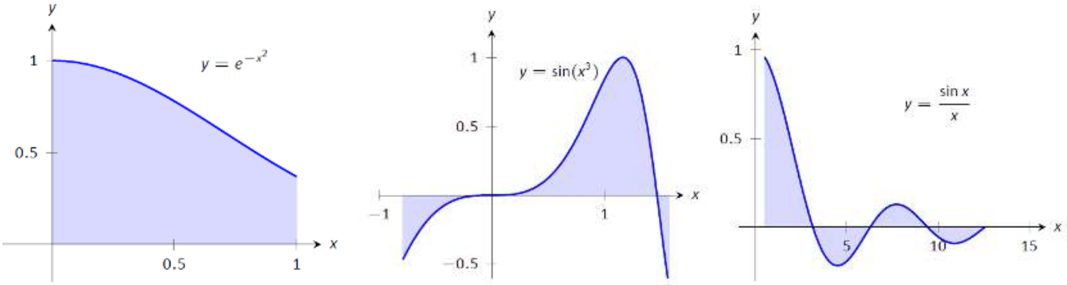


Figure* 5.5.1*: Graphically representing three definite integrals that cannot be evaluated using antiderivatives.

The Left and Right Hand Rule Methods

In Section 5.3 we addressed the problem of evaluating definite integrals by approximating the area under the curve using rectangles. We revisit those ideas here before introducing other methods of approximating definite integrals.

We start with a review of notation. Let f be a continuous function on the interval $[a,b]$. We wish to approximate $\int_a^b f(x)\,dx$. We partition $[a,b]$ into n equally spaced subintervals, each of length $dx = \frac{b-a}{n}$. The endpoints of these subintervals are labeled as

$$x_1=a,\ x_2 = a+dx,\ x_3 = a+ 2dx,\ \ldots,\ x_i = a+(i-1)\ dx,\ \ldots,\ x_{n+1} = b.$$

Key Idea 8 states that to use the Left Hand Rule we use the summation $\sum_{i=1}^n f(x_i)\,dx$ and to use the Right Hand Rule we use $\sum_{i=1}^n f(x_{i+1})\,dx$. We review the use of these rules in the context of examples.

Example 5.5.1: Approximating definite integrals with rectangles

Approximate $\int_0^1 e^{-x^2}\,dx$ using the Left and Right Hand Rules with 5 equally spaced subintervals.

Solution

We begin by partitioning the interval $[0,1]$ into 5 equally spaced intervals. We have $dx = \frac{1-0}{5} = 1/5 = 0.2$, so

$$x_1 = 0,\ x_2 = 0.2,\ x_3 = 0.4,\ x_4 = 0.6,\ x_5 = 0.8,\ \text{and}\ x_6 = 1.$$

Using the Left Hand Rule, we have:

$$\begin{aligned}\sum_{i=1}^{n} f(x_i)\ dx &= \Big(f(x_1)+f(x_2)+f(x_3)+f(x_4)+f(x_5)\Big)\ dx && (5.5.1)\\ &= \Big(f(0)+f(0.2)+f(0.4)+f(0.6)+f(0.8)\Big)\ dx && (5.5.2)\\ &\approx \big(1+0.961+0.852+0.698+0.527)(0.2) && (5.5.3)\\ &\approx 0.808. && (5.5.4)\end{aligned}$$

Using the Right Hand Rule, we have:

$$\begin{aligned}\sum_{i=1}^{n} f(x_{i+1})\ dx &= \Big(f(x_2)+f(x_3)+f(x_4)+f(x_5)+f(x_6)\Big)\ dx && (5.5.5)\\ &= \Big(f(0.2)+f(0.4)+f(0.6)+f(0.8)+f(1)\Big)\ dx && (5.5.6)\\ &\approx \big(0.961+0.852+0.698+0.527+0.368)(0.2) && (5.5.7)\\ &\approx 0.681. && (5.5.8)\end{aligned}$$

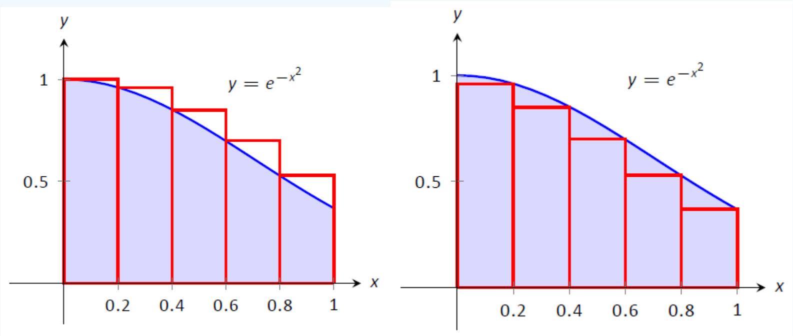


***Figure* 5.5.2*: Approximating $\int_0^1 e^{-x^2}\ dx$ in Example* 5.5.1**

Figure 5.5.2 shows the rectangles used in each method to approximate the definite integral. These graphs show that in this particular case, the Left Hand Rule is an over approximation and the Right Hand Rule is an under approximation. To get a better approximation, we could use more rectangles, as we did in Section 3.1. We could also average the Left and Right Hand Rule results together, giving

$$ \frac{0.808 + 0.681}{2} = 0.7445.\]

The actual answer, accurate to 4 places after the decimal, is 0.7468, showing our average is a good approximation.

Example 5.5.2: Approximating definite integrals with rectangles

Approximate $\int_{-\frac{\pi}{4}}^{\frac{\pi}{2}} \sin(x^3)\ dx$ using the Left and Right Hand Rules with 10 equally spaced subintervals.

Solution

We begin by finding \Delta x:

$$\frac{b-a}{n} = \frac{\pi/2 - (-\pi/4)}{10} = \frac{3\pi}{40}\approx 0.236.\]

It is useful to write out the endpoints of the subintervals in a table; in Table 5.5.1, we give the exact values of the endpoints, their decimal approximations, and decimal approximations of $\sin(x^3)$ evaluated at these points.

Table 5.5.1: Table of values used to approximate $\int_{-\frac{\pi}{4}}^{\frac{\pi}{2}} \sin(x^3)\ dx$ in Example 5.5.1.

x_i	Exact	Approx.	$\sin(x_i^3)$
x_1	$-\pi/4$	-0.785	-0.466
x_2	$-7\pi/40$	-0.550	-0.165
x_3	$-\pi/10$	-0.314	-0.031
x_4	$-\pi/40$	-0.0785	0
x_5	$\pi/20$	0.157	0.004

x_i	Exact	Approx.	$\sin(x_i^3)$
x_6	$\pi/8$	0.393	0.061
x_7	$\pi/5$	0.628	0.246
x_8	$11\pi/40$	0.864	0.601
x_9	$7\pi/20$	1.10	0.971
x_{10}	$17\pi/40$	1.34	0.690
x_{11}	$\pi/2$	1.57	-0.670

Once this table is created, it is straightforward to approximate the definite integral using the Left and Right Hand Rules. (Note: the table itself is easy to create, especially with a standard spreadsheet program on a computer. The last two columns are all that are needed.) The Left Hand Rule sums the first 10 values of $\sin(x_i^3)$ and multiplies the sum by dx; the Right Hand Rule sums the last 10 values of $\sin(x_i^3)$ and multiplies by dx. Therefore we have:

Left Hand Rule: $\int_{-\frac{\pi}{4}}^{\frac{\pi}{2}} \sin(x^3)\, dx \approx (1.91)(0.236) = 0.451.$

Right Hand Rule: $\int_{-\frac{\pi}{4}}^{\frac{\pi}{2}} \sin(x^3)\, dx \approx (1.71)(0.236) = 0.404.$

Average of the Left and Right Hand Rules: 0.4275.

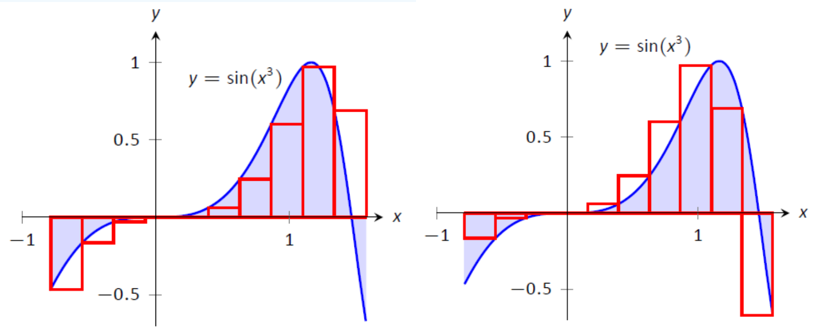


Figure $5.5.3$: $\int_{-\frac{\pi}{4}}^{\frac{\pi}{2}} \sin(x^3)\, dx$ *in Example* $5.5.2$

The actual answer, accurate to 3 places after the decimal, is 0.460. Our approximations were once again fairly good. The rectangles used in each approximation are shown in Figure $5.5.3$. It is clear from the graphs that using more rectangles (and hence, narrower rectangles) should result in a more accurate approximation.

The Trapezoidal Rule

In Example $5.5.1$ we approximated the value of $\int_0^1 e^{-x^2}\, dx$ with 5 rectangles of equal width. Figure $5.5.2$ shows the rectangles used in the Left and Right Hand Rules. These graphs clearly show that rectangles do not match the shape of the graph all that well, and that accurate approximations will only come by using lots of rectangles.

Instead of using rectangles to approximate the area, we can instead use *trapezoids*. In Figure $5.5.4$, we show the region under $f(x) = e^{-x^2}$ on $[0,1]$ approximated with 5 trapezoids of equal width; the top "corners" of each trapezoid lies on the graph of $f(x)$. It is clear from this figure that these trapezoids more accurately approximate the area under f and hence should give a better approximation of $\int_0^1 e^{-x^2}\, dx$. (In fact, these trapezoids seem to give a *great* approximation of the area!)

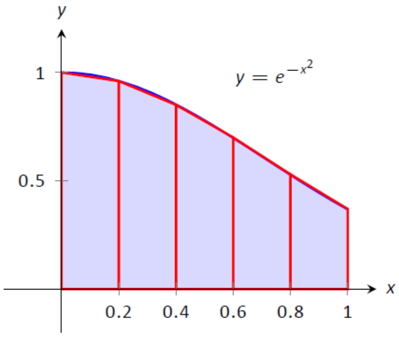


***Figure** 5.5.4: Approximating $\int_0^1 e^{-x^2}\,dx$ using 5 trapezoids of equal widths.*

The formula for the area of a trapezoid is given in Figure 5.5.5. We approximate $\int_0^1 e^{-x^2}\,dx$ with these trapezoids in the following example.

Example 5.5.3: Approximating definite integrals using trapezoids

Use 5 trapezoids of equal width to approximate $\int_0^1 e^{-x^2}\,dx$.

Solution

To compute the areas of the 5 trapezoids in Figure 5.5.6, it will again be useful to create a table of values as shown in Table 5.5.2.

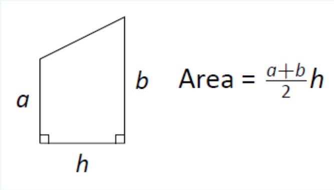


***Figure** 5.5.5: The area of a trapezoid*

Table 5.5.2: A table of values of e^{-x^2}.

x_i	$e^{-x_i^2}$
0	1
0.2	0.961
0.4	0.852
0.6	0.698
0.8	0.527
1	0.368

The leftmost trapezoid has legs of length 1 and 0.961 and a height of 0.2. Thus, by our formula, the area of the leftmost trapezoid is:

$$ \frac{1+0.961}{2}(0.2) = 0.1961.\]

Moving right, the next trapezoid has legs of length 0.961 and 0.852 and a height of 0.2. Thus its area is:

$\frac{0.961+0.852}2(0.2) = 0.1813.\]

The sum of the areas of all 5 trapezoids is:

$$\frac{1+0.961}{2}(0.2)+\frac{0.961+0.852}{2}(0.2)+\frac{0.852+0.698}{2}(0.2)+ \tag{5.5.9}$$

$$\frac{0.698+0.527}{2}(0.2)+\frac{0.527+0.368}{2}(0.2)=0.7445. \tag{5.5.10}$$

We approximate $\int_0^1 e^{-x^2}\,dx \approx 0.7445$.

There are many things to observe in this example. Note how each term in the final summation was multiplied by both 1/2 and by $dx = 0.2$. We can factor these coefficients out, leaving a more concise summation as:

$$\frac12(0.2)\Big[(1+0.961) + (0.961+0.852) + (0.852+0.698) + (0.698+ 0.527) +(0.527 + 0.368)\Big].\]

Now notice that all numbers except for the first and the last are added twice. Therefore we can write the summation even more concisely as

$$\frac{0.2}{2}\Big[1 + 2(0.961+0.852+0.698+0.527) + 0.368\Big].\]

This is the heart of the **Trapezoidal Rule**, wherein a definite integral $\int_a^b f(x)\,dx$ is approximated by using trapezoids of equal widths to approximate the corresponding area under f. Using n equally spaced subintervals with endpoints $x_1, x_2, \ldots, x_{n+1}$, we again have $\Delta x = \frac{b-a}{n}$. Thus:

$$\int_a^b f(x)\,dx \approx \sum_{i=1}^{n} \frac{f(x_i)+f(x_{i+1})}{2}\,dx \tag{5.5.11}$$

$$= \frac{dx}{2}\sum_{i=1}^{n}\big(f(x_i)+f(x_{i+1})\big) \tag{5.5.12}$$

$$= \frac{dx}{2}\Big[f(x_1) + 2\sum_{i=2}^{n} f(x_i) + f(x_{n+1})\Big]. \tag{5.5.13}$$

Example 5.5.4: Using the Trapezoidal Rule

Revisit Example 5.5.4 and approximate $\int_{-\frac{\pi}{4}}^{\frac{\pi}{2}} \sin(x^3)\,dx$ using the Trapezoidal Rule and 10 equally spaced subintervals.

Solution

We refer back to Table 5.5.1 for the table of values of $\sin(x^3)$. Recall that $dx = 3\pi/40 \approx 0.236$. Thus we have:

$$\int_{-\frac{\pi}{4}}^{\frac{\pi}{2}} \sin(x^3)\,dx \approx \frac{0.236}{2}\Big[-0.466 + 2\Big(-0.165 + (-0.031) + \ldots + 0.69\Big) + (-0.67)\Big]$$
$$= 0.4275.$$

Notice how "quickly" the Trapezoidal Rule can be implemented once the table of values is created. This is true for all the methods explored in this section; the real work is creating a table of x_i and $f(x_i)$ values. Once this is completed, approximating the definite integral is not difficult. Again, using technology is wise. Spreadsheets can make quick work of these computations and make using lots of subintervals easy.

Also notice the approximations the Trapezoidal Rule gives. It is the average of the approximations given by the Left and Right Hand Rules! This effectively renders the Left and Right Hand Rules obsolete. They are useful when first learning about definite integrals, but if a real approximation is needed, one is generally better off using the Trapezoidal Rule instead of either the Left or Right Hand Rule.

How can we improve on the Trapezoidal Rule, apart from using more and more trapezoids? The answer is clear once we look back and consider what we have *really* done so far. The Left Hand Rule is not *really* about using rectangles to approximate area. Instead, it approximates a function f with constant functions on small subintervals and then computes the definite integral of these constant functions. The Trapezoidal Rule is really approximating a function f with a linear function on a small subinterval, then computes the definite integral of this linear function. In both of these cases the definite integrals are easy to compute in geometric terms.

So we have a progression: we start by approximating f with a constant function and then with a linear function. What is next? A quadratic function. By approximating the curve of a function with lots of parabolas, we generally get an even better approximation of the definite integral. We call this process **Simpson's Rule**, named after Thomas Simpson (1710-1761), even though others had used this rule as much as 100 years prior.

Simpson's Rule

Given one point, we can create a constant function that goes through that point. Given two points, we can create a linear function that goes through those points. Given three points, we can create a quadratic function that goes through those three points (given that no two have the same x--value).

Consider three points (x_1, y_1), (x_2, y_2) and (x_3, y_3) whose x--values are equally spaced and $x_1 < x_2 < x_3$. Let f be the quadratic function that goes through these three points. It is not hard to show that

$$\int_{x_1}^{x_3} f(x)\ dx = \frac{x_3 - x_1}{6}\big(y_1 + 4y_2 + y_3\big). \tag{5.5.14}$$

Consider Figure $5.5.6$. A function f goes through the 3 points shown and the parabola g that also goes through those points is graphed with a dashed line. Using our equation from above, we know exactly that

$$\int_1^3 g(x)\ dx = \frac{3-1}{6}\big(3 + 4(1) + 2\big) = 3. \tag{5.5.15}$$

Since g is a good approximation for f on $[1, 3]$, we can state that $$\int_1^3 f(x)\ dx \approx 3.\]

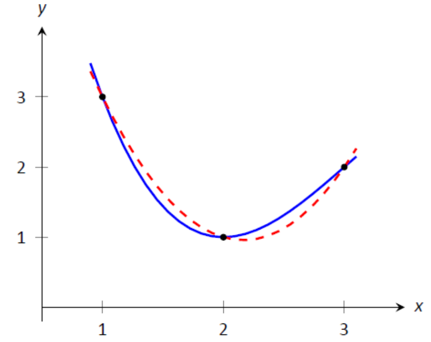


***Figure** $5.5.6$**:** A graph of a function f and a parabola that approximates it well on $[1, 3]$.*

Notice how the interval $[1, 3]$ was split into two subintervals as we needed 3 points. Because of this, whenever we use Simpson's Rule, we need to break the interval into an even number of subintervals.

In general, to approximate $\int_a^b f(x)\ dx$ using Simpson's Rule, subdivide $[a, b]$ into n subintervals, where n is even and each subinterval has width $dx = (b-a)/n$. We approximate f with $n/2$ parabolic curves, using Equation $5.5.22$ to compute the area under these parabolas. Adding up these areas gives the formula:

$$\int_a^b f(x) \ dx \approx \frac{\ dx}3\Big[f(x_1)+4f(x_2)+2f(x_3)+4f(x_4)+\ldots+2f(x_{n-1})+4f(x_n)+f(x_{n+1})\Big].\]

Note how the coefficients of the terms in the summation have the pattern 1, 4, 2, 4, 2, 4, ..., 2, 4, 1.

Let's demonstrate Simpson's Rule with a concrete example.

Example $5.5.5$: Using Simpson's Rule

Approximate $\int_0^1 e^{-x^2}\ dx$ using Simpson's Rule and 4 equally spaced subintervals.

Solution

We begin by making a table of values as we have in the past, as shown in Table $5.5.3$.

Table $5.5.3$: A table of values to approximate $\int_0^1 e^{-x^2}\ dx$, along with a graph of the function.

x_i	$e^{-x_i^2}$
0	1
0.25	0.939
0.5	0.779
0.75	0.570
1	0.368

Simpson's Rule states that

`$$\int_0^1e^{-x^2}\ dx \approx \frac{0.25}{3}\Big[1+4(0.939)+2(0.779)+4(0.570) + 0.368\Big] = 0.7468\overline{3}.\]`

Recall in Example $5.5.1$ we stated that the correct answer, accurate to 4 places after the decimal, was 0.7468. Our approximation with Simpson's Rule, with 4 subintervals, is better than our approximation with the Trapezoidal Rule using 5!

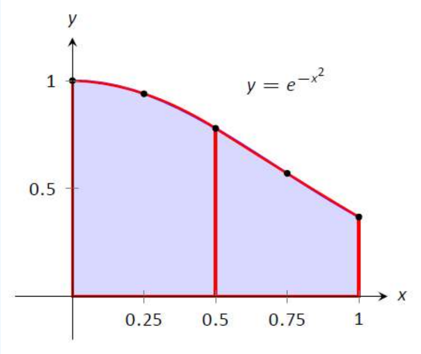


***Figure** $5.5.7$**: Using Simpson's Rule with $n = 4$ to approximate $\int_0^1 e^{-x^2}\,dx$.*

Figure $5.5.7$ shows $f(x) = e^{-x^2}$ along with its approximating parabolas, demonstrating how good our approximation is. The approximating curves are nearly indistinguishable from the actual function.

Example $5.5.6$: Using Simpson's Rule

Approximate $\int_{-\frac{\pi}{4}}^{\frac{\pi}{2}} \sin(x^3)\,dx$ using Simpson's Rule and 10 equally spaced intervals.

Solution

Table $5.5.4$ shows the table of values that we used in the past for this problem, shown here again for convenience. Again, $dx = (\pi/2 + \pi/4)/10 \approx 0.236$.

Table $5.5.4$: Table of values used to approximate $\int_{-\frac{\pi}{4}}^{\frac{\pi}{2}} \sin(x^3)\,dx$ in Example $5.5.6$.

x_i	$\sin(x_i^3)$
-0.785	-0.466
-0.550	-0.165
-0.314	-0.031
-0.0785	0
0.157	0.004
0.393	0.061
0.628	0.246
0.864	0.601
1.10	0.971
1.34	0.690
1.57	-0.670

Simpson's Rule states that

$$\int_{-\frac{\pi}{4}}^{\frac{\pi}{2}} \sin(x^3)\,dx \approx \frac{0.236}{3}\Big[(-0.466) + 4(-0.165) + 2(-0.031) + \ldots \tag{5.5.16}$$

$$\ldots + 2(0.971) + 4(0.69) + (-0.67)\Big] \tag{5.5.17}$$

$$= 0.4701 \tag{5.5.18}$$

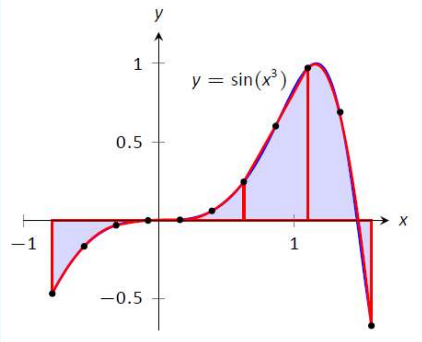


***Figure** 5.5.8: Approximating* $\int_{-\frac{\pi}{4}}^{\frac{\pi}{2}} \sin(x^3)\ dx$ *in Example* 5.5.6 *with Simpson's Rule and 10 equally spaced intervals.*

Recall that the actual value, accurate to 3 decimal places, is 0.460. Our approximation is within one 1/100th of the correct value. The graph in Figure 5.5.8 shows how closely the parabolas match the shape of the graph.

Summary and Error Analysis

We summarize the key concepts of this section thus far in the following Key Idea.

Key Idea 9: Numerical Integration

Let f be a continuous function on $[a,b]$, let n be a positive integer, and let $\Delta x = \frac{b-a}{n}$.

Set $x_1 = a$, $x_2 = a +\ dx$, ..., $x_i = a + (i-1)\ dx$, $x_{n+1} = b$.

Consider $\int_a^b f(x)\ dx$.

Left Hand Rule: $\int_a^b f(x)\ dx \approx\ dx\Big[f(x_1) + f(x_2) + \ldots + f(x_n)\Big]$.

Right Hand Rule: $\int_a^b f(x)\ dx \approx\ dx\Big[f(x_2) + f(x_3) + \ldots + f(x_{n+1})\Big]$.

Trapezoidal Rule: $\int_a^b f(x)\ dx \approx \frac{dx}{2}\Big[f(x_1) + 2f(x_2) + 2f(x_3) + \ldots + 2f(x_n) + f(x_{n+1})\Big]$.

Simpson's Rule: $\int_a^b f(x)\ dx \approx \frac{dx}{3}\Big[f(x_1) + 4f(x_2) + 2f(x_3) + \ldots + 4f(x_n) + f(x_{n+1})\Big]$ (n even)}.

In our examples, we approximated the value of a definite integral using a given method then compared it to the "right" answer. This should have raised several questions in the reader's mind, such as:

1. How was the "right" answer computed?
2. If the right answer can be found, what is the point of approximating?
3. If there is value to approximating, how are we supposed to know if the approximation is any good?

These are good questions, and their answers are educational. In the examples, *the* right answer was never computed. Rather, an approximation accurate to a certain number of places after the decimal was given. In Example 5.5.1, we do not know the *exact* answer, but we know it starts with 0.7468. These more accurate approximations were computed using numerical integration but with more precision (i.e., more subintervals and the help of a computer).

Since the exact answer cannot be found, approximation still has its place. How are we to tell if the approximation is any good?

"Trial and error" provides one way. Using technology, make an approximation with, say, 10, 100, and 200 subintervals. This likely will not take much time at all, and a trend should emerge. If a trend does not emerge, try using yet more subintervals. Keep in mind that trial and error is never foolproof; you might stumble upon a problem in which a trend will not emerge.

A second method is to use Error Analysis. While the details are beyond the scope of this text, there are some formulas that give *bounds* for how good your approximation will be. For instance, the formula might state that the approximation is within 0.1 of the correct answer. If the approximation is 1.58, then one knows that the correct answer is between 1.48 and 1.68. By using lots of subintervals, one can get an approximation as accurate as one likes. Theorem 5.5.1 states what these bounds are.

Theorem 5.5.1: Error Bounds in the Trapezoidal and Simpson's Rules

1. Let E_T be the error in approximating $\int_a^b f(x)\ dx$ using the Trapezoidal Rule.
   If f has a continuous 2nd derivative on $[a,b]$ and M is any upper bound of $|f''(x)|$ on $[a,b]$, then

$$E_T \leq \frac{(b-a)^3}{12n^2}M.$$

2. Let E_S be the error in approximating $\int_a^b f(x)\ dx$ using Simpson's Rule.
   If f has a continuous 4th derivative on $[a,b]$ and M is any upper bound of $|f^{(4)}|$ on $[a,b]$, then

$$E_S \leq \frac{(b-a)^5}{180n^4}M.$$

There are some key things to note about this theorem.

1. The larger the interval, the larger the error. This should make sense intuitively.
2. The error shrinks as more subintervals are used (i.e., as n gets larger).
3. The error in Simpson's Rule has a term relating to the 4th derivative of f. Consider a cubic polynomial: it's 4^{th} derivative is 0. Therefore, the error in approximating the definite integral of a cubic polynomial with Simpson's Rule is 0 -- Simpson's Rule computes the exact answer!

We revisit Examples 5.5.3 and 5.5.5 and compute the error bounds using Theorem 5.5.1 in the following example.

Example 5.5.7: Computing error bounds

Find the error bounds when approximating $\int_0^1 e^{-x^2}\ dx$ using the Trapezoidal Rule and 5 subintervals, and using Simpson's Rule with 4 subintervals.

Solution

Trapezoidal Rule with $n=5$:

We start by computing the 2^{nd} derivative of $f(x)=e^{-x^2}$:

$$f''(x)=e^{-x^2}(4x^2-2). \tag{5.5.19}$$

Figure 5.5.8 shows a graph of $f''(x)$ on $[0,1]$. It is clear that the largest value of f'', in absolute value, is 2. Thus we let $M=2$ and apply the error formula from Theorem 5.5.1.

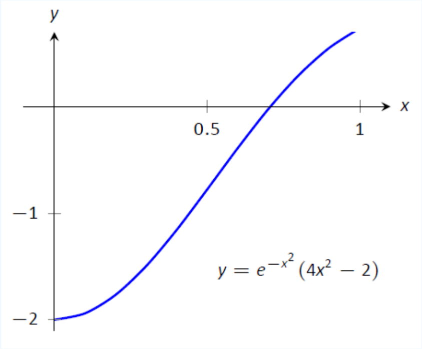


***Figure** 5.5.9: Graphing $f''(x)$ in Example 5.5.7 to help establish error bounds.*

$$E_T = \frac{(1-0)^3}{12\cdot 5^2}\cdot 2 = 0.00\overline{6}.$$

Our error estimation formula states that our approximation of 0.7445 found in Example 5.5.3 is within 0.0067 of the correct answer, hence we know that

$$0.7445-0.0067 = .7378 \leq \int_0^1e^{-x^2}\ dx \leq 0.7512 = 0.7445 + 0.0067.$$

We had earlier computed the exact answer, correct to 4 decimal places, to be 0.7468, affirming the validity of Theorem 5.5.1.

Simpson's Rule with $n=4$:

We start by computing the 4^{th} derivative of $f(x)=e^{-x^2}$:

$$f\,^{(4)}(x) = e^{-x^2}(16x^4-48x^2+12).$$

Figure $5.5.9$ shows a graph of $f^{(4)}(x)$ on $[0,1]$. It is clear that the largest value of $f^{(4)}$, in absolute value, is 12. Thus we let $M = 12$ and apply the error formula from Theorem $5.5.1$.

$$E_s = \frac{(1-0)^5}{180\cdot 4^4}\cdot 12 = 0.00026.$$

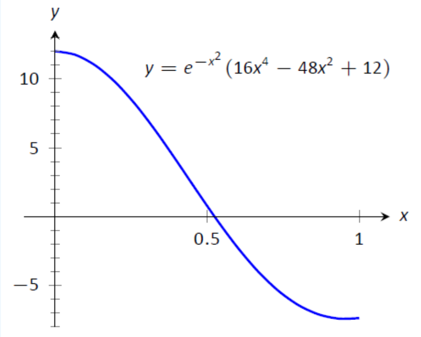


***Figure** $5.5.10$**:** Graphing $f^{(4)}(x)$ in Example $5.5.7$ to help establish error bounds.*

Our error estimation formula states that our approximation of $0.7468\overline{3}$ found in Example $5.5.5$ is within 0.00026 of the correct answer, hence we know that

$$0.74683-0.00026 = .74657 \leq \int_0^1e^{-x^2}\ dx \leq 0.74709 = 0.74683 + 0.00026.$$

Once again we affirm the validity of Theorem$5.5.1$.

At the beginning of this section we mentioned two main situations where numerical integration was desirable. We have considered the case where an antiderivative of the integrand cannot be computed. We now investigate the situation where the integrand is not known. This is, in fact, the most widely used application of Numerical Integration methods. "Most of the time" we observe behavior but do not know "the" function that describes it. We instead collect data about the behavior and make approximations based off of this data. We demonstrate this in an example.

Example $5.5.8$: Approximating distance traveled

One of the authors drove his daughter home from school while she recorded their speed every 30 seconds. The data is given in Table $5.5.5$. Approximate the distance they traveled.

Solution

Recall that by integrating a speed function we get distance traveled. We have information about $v(t)$; we will use Simpson's Rule to approximate $\int_a^b v(t)\ dt$.

The most difficult aspect of this problem is converting the given data into the form we need it to be in. The speed is measured in miles per hour, whereas the time is measured in 30 second increments.

Table $5.5.5$: Speed data collected at 30 second intervals for Example $5.5.8$.

0	0
1	25
2	22
3	19
4	39
5	0
6	43
7	59
8	54

9	51
10	43
11	35
12	40
13	43
14	30
15	0
16	0
17	28
18	40
19	42
20	40
21	39
22	40
23	23
24	0

We need to compute $dx = (b-a)/n$. Clearly, $n = 24$. What are a and b? Since we start at time $t = 0$, we have that $a = 0$. The final recorded time came after 24 periods of 30 seconds, which is 12 minutes or 1/5 of an hour. Thus we have

$$\ dx = \frac{b-a}{n} = \frac{1/5-0}{24} = \frac1{120}; \quad \frac{\ dx}{3} = \frac{1}{360}.\]

Thus the distance traveled is approximately:

$$\begin{aligned}\int_0^{0.2} v(t)\ dt &\approx \frac{1}{360}\Big[f(x_1)+4f(x_2)+2f(x_3)+\cdots+4f(x_n)+f(x_{n+1})\Big]\\ &= \frac{1}{360}\Big[0+4\cdot 25+2\cdot 22+\cdots+2\cdot 40+4\cdot 23+0\Big]\\ &\approx 6.2167 \text{ miles.}\end{aligned}$$

We approximate the author drove 6.2 miles. (Because we are sure the reader wants to know, the author's odometer recorded the distance as about 6.05 miles.)

We started this chapter learning about antiderivatives and indefinite integrals. We then seemed to change focus by looking at areas between the graph of a function and the x-axis. We defined these areas as the definite integral of the function, using a notation very similar to the notation of the indefinite integral. The Fundamental Theorem of Calculus tied these two seemingly separate concepts together: we can find areas under a curve, i.e., we can evaluate a definite integral, using antiderivatives.

We ended the chapter by noting that antiderivatives are sometimes more than difficult to find: they are impossible. Therefore we developed numerical techniques that gave us good approximations of definite integrals.

We used the definite integral to compute areas, and also to compute displacements and distances traveled. There is far more we can do than that. In Chapter 7 we'll see more applications of the definite integral. Before that, in Chapter 6 we'll learn advanced techniques of integration, analogous to learning rules like the Product, Quotient and Chain Rules of differentiation.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

- Integrated by Justin Marshall.

This page titled 5.5: Numerical Integration is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

5.E: Applications of Integration (Exercises)

5.1: Antiderivatives and Indefinite Integration

Terms and Concepts

1. Define the term “antiderivative” in your own words

2. Is it more accurate to refer to “the” antiderivative of $f(x)$ or “an” antiderivative of $f(x)$?

3. Use your own words to define the indefinite integral of $f(x)$.

4. Fill in the blanks: “Inverse operations do the ____ things in the _____ order.”

5. What is an “initial value problem”?

6. The derivative of a position function is a _____ function.

7. The antiderivative of an acceleration function is a ______ function.

Problems

In Exercises 8-26, evaluate the given indefinite integral.

8. $\int 3x^3\,dx$

9. $\int x^8\,dx$

10. $\int (10x^2-2)\,dx$

11. $\int\ dt$

12. $\int 1\,ds$

13. $\int \frac{1}{3t^2}\,dt$

14. $\int \frac{1}{t^2}\,dt$

15. $\int \frac{1}{\sqrt{x}}\,dx$

16. $\int \sec^2\theta\,d\theta$

17. $\int \sin\theta\,d\theta$

18. $\int (\sec x\tan x+\csc x\cot x)\,dx$

19. $\int 5e^\theta\,d\theta$

20. $\int 3^t\,dt$

21. $\int \frac{5^t}{2}\,dt$

22. $\int (2t+3)^2\,dt$

23. $\int (t^2+3)(t^3-2t)\,dt$

24. $\int x^2x^3\,dx$

25. $\int e^\pi\,dx$

26. $\int a\,dx$

27. This problem investigates why Theorem 35 states that $\int \frac{1}{x}\,dx=\ln|x|+C$.
(a) What is the domain of $y=\ln x$?
(b) Find $\frac{d}{dx}(\ln x)$.
(c) What is the domain of $y=\ln(-x)$?
(d) Find $\frac{d}{dx}\left(\ln(-x)\right)$.
(e) You should find that $1/x$ has two types of antiderivatives, depending on whether $x>0$ or $x<0$. In one expression, give a formula for $\int \frac{1}{x}\,dx$ that takes these different domains into account, and explain your answer.

In Exercises 28-38, find $f(x)$ described by the given initial value problem.

28. $f'(x) = \sin x$ and $f(0) = 2$

29. $f'(x) = 5e^x$ and $f(0) = 10$

30. $f'(x) = 4x^3 - 3x^2$ and $f(-1) = 9$

31. $f'(x) = \sec^2 x$ and $f(\pi/4) = 5$

32. $f'(x) = 7^x$ and $f(2) = 1$

33. $f''(x) = 5$ and $f'(0) = 7, f(0) = 3$

34. $f''(x) = 7x$ and $f'(1) = -1, f(1) = 10$

35. $f''(x) = 5e^x$ and $f'(0) = 3, f(0) = 5$

36. $f''(\theta) = \sin\theta$ and $f'(\pi) = 2, f(\pi) = 4$

37. $f''(x) = 24x^2 + 2^x - \cos x$ and $f'(0) = 5, f(0) = 0$

38. $f''(x) = 0$ and $f'(1) = 3, f(1) = 1$

Review

39. Use information gained from the first and second derivative to sketch $f(x) = \frac{1}{e^x+1}$.

40. Given $y = x^2 e^x \cos x$, find dy.

5.2: The Definite Integral

Terms and Concepts

1. What is "total signed area"?

2. What is "displacement"?

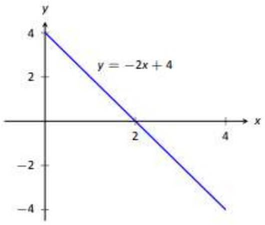
3. What is $\int_3^3 \sin x \, dx$

4. Give a single definite integral that has the same value as $\int_0^1 (2x+3)\,dx + \int_1^2 (2x+3)\,dx$.

Problems

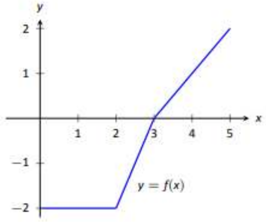
In Exercises 5-9, a graph of a function $f(x)$ is given. Using the geometry of the graph, evaluate the definite integrals.

5.



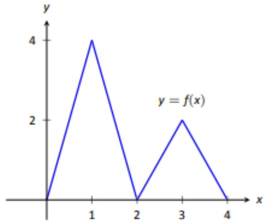
(a) $\int_0^1 (-2x+4)\,dx$
(b) $\int_0^2 (-2x+4)\,dx$
(c) $\int_0^3 (-2x+4)\,dx$
(d) $\int_1^3 (-2x+4)\,dx$
(e) $\int_2^4 (-2x+4)\,dx$
(f) $\int_0^1 (-6x+12)\,dx$

6.



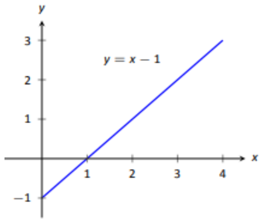
(a) $\int_0^2 f(x)\,dx$
(b) $\int_0^3 f(x)\,dx$
(c) $\int_0^5 f(x)\,dx$
(d) $\int_2^5 f(x)\,dx$
(e) $\int_5^3 f(x)\,dx$
(f) $\int_0^3 f(x)\,dx$

7.



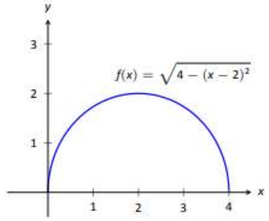
(a) $\int_0^2 f(x)\,dx$
(b) $\int_2^4 f(x)\,dx$
(c) $\int_2^4 2f(x)\,dx$
(d) $\int_0^1 4x\,dx$
(e) $\int_2^3 (2x-4)\,dx$
(f) $\int_2^3 (4x-8)\,dx$

8.



(a) $\int_0^1 (x-1)\,dx$
(b) $\int_0^2 (x-1)\,dx$
(c) $\int_0^3 (x-1)\,dx$
(d) $\int_2^3 (x-1)\,dx$
(e) $\int_1^4 (x-1)\,dx$
(f) $\int_1^4 \left((x-1)+1\right)\,dx$

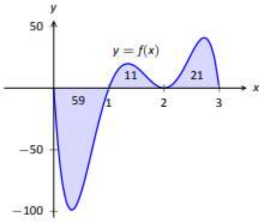
9.



(a) $\int_0^2 f(x)\,dx$
(b) $\int_2^4 f(x)\,dx$
(c) $\int_0^4 f(x)\,dx$
(d) $\int_0^4 5f(x)\,dx$

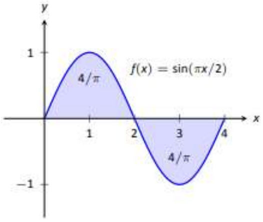
In Exercises 10-13, a graph of a function $f(x)$ is given; the numbers inside the shaded regions give the area of that region. Evaluate the definite integrals using this area information.

10.



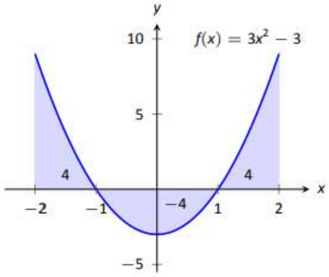
(a) $\int_0^1 f(x)\,dx$
(b) $\int_0^2 f(x)\,dx$
(c) $\int_0^3 f(x)\,dx$
(d) $\int_1^2 -3f(x)\,dx$

11.



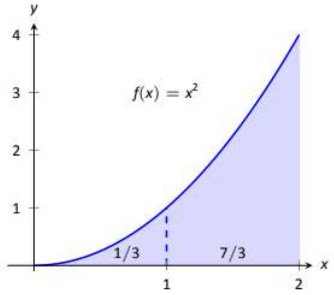
(a) $\int_0^2 f(x)\,dx$
(b) $\int_2^4 f(x)\,dx$
(c) $\int_0^4 f(x)\,dx$
(d) $\int_0^1 f(x)\,dx$

12.



(a) $\int_{-2}^{-1} f(x)\,dx$
(b) $\int_1^2 f(x)\,dx$
(c) $\int_{-1}^1 f(x)\,dx$
(d) $\int_0^1 f(x)\,dx$

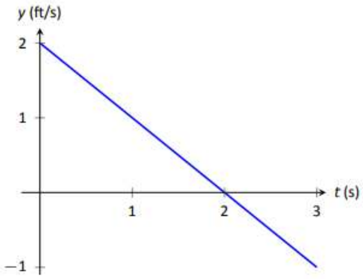
13.



(a) $\int_0^2 5x^2\,dx$
(b) $\int_0^2 (x^2+1)\,dx$
(c) $\int_1^3 (x-1)^2\,dx$
(d) $\int_2^4 ((x-2)+5)\ dx$

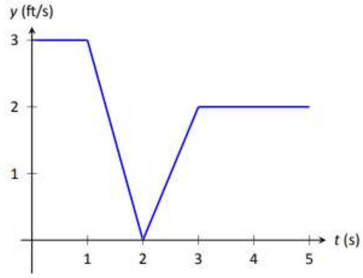
In Exercises 14-15, a graph of the velocity function of an object moving in a straight line is given. Answer the questions based on that graph.

14.



(a) What is the object's maximum velocity?
(b) What is the object's maximum displacement?
(c) What is the object's total displacement on [0,3]?

15.



(a) What is the object's maximum velocity?
(b) What is the object's maximum displacement?
(c) What is the object's total displacement on [0,5]?

16. An object is thrown straight up with a velocity, in ft/s, given by $v(t) = -32t + 64$, where t is in seconds, from a height of 48 feet.
(a) What is the object's maximum velocity?
(b) What is the object's maximum displacement?
(c) When does the maximum displacement occur?
(d) When will the object reach a height of 0? (Hint: find when the displacement is -48ft.)

17. An object is thrown straight up with a velocity, in ft/s, given by $v(t) = -32t + 96$, where t is seconds, from a height of 64 feet.
(a) What is the object's initial velocity?
(b) What is the object's displacement 0?

(c) How long does it take for the object to return to its initial height?
(d) When will the object reach a height of 210ft?

In Exercises 18-21, let

- $\int_0^2 f(x)\,dx = 5,$
- $\int_0^3 f(x)\,dx = 7,$
- $\int_0^2 g(x)\,dx = -3$, and
- $\int_2^3 g(x)\,dx = 5.$

Use these values to evaluate the given definite integrals.

18. $\int_0^2 (f(x)+g(x))\ dx$

19. $\int_0^3 (f(x)-g(x))\ dx$

20. $\int_2^3 (3f(x)+2g(x))\ dx$

21. Find values for a and b such that
$\int_0^3 (af(x)+bg(x))\ dx = 0$

In Exercises 22-25, let

- $\int_0^3 s(t)\ dt = 10,$
- $\int_3^5 s(t)\ dt = 8,$
- $\int_3^5 r(t)\,dt = -1$, and
- $\int_0^5 r(t)\,dt = 11.$

Use these values to evaluate the given definite integrals.

22. $\int_0^3 (s(t)+r(t))\ dt$

23. $\int_5^0 (s(t)-r(t))\ dt$

24. $\int_3^3 (\pi s(t)-7r(t))\ dt$

25. Find values for *a* and *b* such that
$\int_0^5 (ar(t)+bs(t))\ dt = 0$

Review

In Exercises 26-29, evaluate the given indefinite integral.

26. $\int (x^3-2x^2+7x-9)\,dx$

27. $\int (\sin x - \cos x + \sec^2 x)\,dx$

28. $\int (\sqrt[3]{t} + \frac{1}{t^2} + 2^t)\,dt$

29. $\int \left(\frac{1}{x} - \csc x \cot x\right)\ dx$

5.3: Riemann Sums

Terms and Concepts

1. A fundamental calculus technique is to use ________ to refine approximations to get an exact answer.

2. What is the upper bound in the summation $\sum_{i=7}^{14}(48i-201)$?

3. This section approximates definite integrals using what geometric shape?

4. T/F: A sum using the Right Hand Rule is an example of a Riemann Sum.

Problems

In Exercises 5-11, write out each term of the summation and compute the sum.

5. $\sum_{i=2}^{4} i^2$

6. $\sum_{i=-1}^{3}(4i-2)$

7. $\sum_{i=-2}^{2}\sin(\pi i/2)$

8. $\sum_{i=1}^{5}\frac{1}{i}$

9. $\sum_{i=1}^{6}(-1)^i i$

10. $\sum_{i=1}^{4}\left(\frac{1}{i}-\frac{1}{i+1}\right)$

11. $\sum_{i=0}^{5}(-1)^i\cos(\pi i)$

In Exercises 12-15, write each sum in summation notation.

12. $3+6+9+12+15$

13. $-1+0+3+8+15+24+35+48+63$

14. $\frac{1}{2}+\frac{2}{3}+\frac{3}{4}+\frac{4}{5}$

15. $1-e+e^2-e^3+e^4$

In Exercises 16-22, evaluate the summation using Theorem 37.

16. $\sum_{i=1}^{2} 5i$

17. $\sum_{i=1}^{1} 0(3i^2-2i)$

18. $\sum_{i=1}^{1} 5(2i^3-10)$

19. $\sum_{i=1}^{1} 0(-4i^3+10i^2-7i+11)$

20. $\sum_{i=1}^{1} 0(i^3-3t^2+2i+7)$

21. $1+2+3+...+99+100$

22. $1+4+9+...+361+400$

Theorem 37 states

$\sum_{i=1}^{n} a_i=\sum+i=1^k a_i+\sum_{i=k+1}^{n} a_i$, so

$\sum_{i=k+1}^{n} a_i=\sum_{i=1}^{n} a_i-\sum_{i=1}^{k} a_i$.

Use this fact, along with other parts of Theorem 37, to evaluate the summations given in Exercises 23-26.

23. $\sum_{i=11}^{2} 0i$

24. $\sum_{i=16}^{2} 5i^3$

25. $\sum_{i=7}^{1} 24$

26. $\sum_{i=5}^{1} 04i^3$

In Exercises 27-32, a definite integral
$\int_a^b f(x)\,dx$ **is given.**
(a) Graph $f(x)$ **on** [a,b].
(b) Add to the sketch rectangles using the provided rule.
(c) Approximate $\int_a^b f(x)\,dx$ **by summing the areas of the rectangles.**

27. $\int_{-3}^{3} x^2\,dx$, with 6 rectangles using the Left Hand Rule.

28. $\int_0^2 (5-x^2)\,dx$, with 4 rectangles using the Midpoint Rule.

29. $\int_0^{\pi}\sin x\,dx$, with 6 rectangles using the Right Hand Rule.

30. $\int_0^3 2^x\,dx$, with 5 rectangles using the Left Hand Rule.

31. $\int_1^2 \ln x\,dx$, with 3 rectangles using the Midpoint Rule.

32. $\int_1^9 \frac{1}{x}\,dx$, with 4 rectangles using the Right Hand Rule.

In Exercises 33-38, a definite integral
$\int_a^b f(x)\,dx$ **is given. As demonstrated in Example 123 and 124, do the following.**
(a) Find a formula to approximate $\int_a^b f(x)\,dx$ **using** n **subintervals and the provided rule.**

(b) Evaluate the formula using $n = 10,\ 100$ and $1000.$

(c) Find the limit of the formula, as $n \to \infty$**, to find the exact value of** $\int_a^b f(x)\,dx$**.**

33. $\int_0^1 x^3\,dx$, using the Right Hand Rule.

34. $\int_{-1}^1 3x^2\,dx$, using the Left Hand Rule.

35. $\int_{-1}^3 (3x-1)\,dx$, using the Midpoint Rule.

36. $\int_1^4 (2x^2-3)\,dx$, using the Left Hand Rule.

37. $\int_{-10}^{10} (5-x)\,dx$, using the Right Hand Rule.

38. $\int_0^1 (x^3-x^2)\,dx$, using the Right Hand Rule.

Review

In Exercises 39-44, find an antiderivative of the given function.

39. $f(x) = 5\sec^2 x$

40. $f(x) = \frac{7}{x}$

41. $g(t) = 4t^5 - 5t^3 + 8$

42. $g(t) = 5 \cdot 8^t$

43. $g(t) = \cos t + \sin t$

44. $f(x) = \frac{1}{\sqrt{x}}$

5.4: The Fundamental Theorem of Calculus

Terms and Concepts

1. How are the definite and indefinite integrals related?

2. What constants of integration is most commonly used when evaluating definite integrals?

3. T/F: If f is a continuous function, then $F(x) = \int_a^x f(t)\,dt$ is also a continuous function.

4. The definite integral can be used to find "the area under a curve." Give two other uses for definite integrals.

Problems

In Exercises 5-28, evaluate the definite integral.

5. $\int_1^3 (3x^2 - 2x + 1)\,dx$

6. $\int_0^4 (x-1)^2\,dx$

7. $\int_{-1}^1 (x^3 - x^5)\,dx$

8. $\int_{\pi/2}^{\pi} \cos x\,dx$

9. $\int_0^{\pi/4} \sec^2 x\,dx$

10. $\int_1^e \frac{1}{x}\,dx$

11. $\int_{-1}^1 5^x\,dx$

12. $\int_{-2}^{-1} (4 - 2x^3)\,dx$

13. $\int_0^{\pi} (2\cos x - 2\sin x)\,dx$

14. $\int_1^3 e^x\,dx$

15. $\int_0^4 \sqrt{t}\,dt$

16. $\int_9^{25} \frac{1}{\sqrt{t}}\,dt$

17. $\int_1^8 \sqrt[3]{x}\,dx$

18. $\int_1^2 \frac{1}{x}\,dx$

19. $\int_{1}^{2} \frac{1}{x^2}\, dx$

20. $\int_{1}^{2} \frac{1}{x^3}\, dx$

21. $\int_{0}^{1} x\, dx$

22. $\int_{0}^{1} x^2\, dx$

23. $\int_{0}^{1} x^3\, dx$

24. $\int_{0}^{1} x^{100}\, dx$

25. $\int_{-4}^{4} dx$

26. $\int_{-10}^{-5} 3\, dx$

27. $\int_{-2}^{2} 0\, dx$

28. $\int_{\pi/6}^{\pi/3} \csc x \cot x\, dx$

29. Explain why:
(a) $\int_{-1}^{1} x^n\, dx = 0$, when *n* is a positive, odd integer, and
(b) $\int_{-1}^{1} x^n\, dx = 2\int_{0}^{1} x^n\, dx$ when *n* is a positive, even integer.

In Exercises 30-33, find a value c guaranteed by the Mean Value Theorem.

30. $\int_{0}^{2} x^2\, dx$

31. $\int_{-2}^{2} x^2\, dx$

32. $\int_{0}^{1} e^x\, dx$

33 $\int_{0}^{1} 6\sqrt{x}\, dx$

In Exercises 34-39, find the average value of the function on the given interval.

34. $f(x) = \sin x \text{ on } [0, \pi/2]$

35. $y = \sin x \text{ on } [0, \pi]$

36. $y = x \text{ on } [0, 4]$

37. $y = x^2 \text{ on } [0, 4]$

38. $y = x^3 \text{ on } [0, 4]$

39. $g(t) = 1/t \text{ on } [1, e]$

In Exercises 40-44, a velocity function of an object moving along a straight line is given. Find the displacement of the object over the given time interval.

40. $v(t) = -32t + 20$ft/s on [0,5].

41. $v(t) = -32t + 200$ft/s on [0,10].

42. $v(t) = 2^t$mph on [-1,1].

43. $v(t) = \cos t$ft/s on $[0, 3\pi/2]$.

44. $v(t) = \sqrt[4]{t}$ft/s on [0,16].

In Exercises 45-48, an acceleration function of an object moving along a straight line is given. Find the change of the object's velocity over the given time interval.

45. $a(t) = -32$ft/s on [0,2].

46. $a(t) = 10$ft/s on [0,5].

47. $a(t) = t$ft/s^2 on [0,2].

48. $a(t) = \cos t$ft/s^2 on $[0, \pi]$.

In Exercises 49-52, sketch the given functions and fine the area of the enclosed region.

49. $y = 2x,\ y = 5x,\ \text{and}\ x = 3.$

50. $y = -x + 1,\ y = 3x + 6,\ x = 2$ and $x = -1.$

51. $y = x^2 - 2x + 5,\ y = 5x - 5.$

52. $y = 2x^2 + 2x - 5,\ y = x^2 + 3x + 7.$

In Exercises 53-56, find $F'(x)$.

53. $F(x) = \int_2^{x^3+x} \frac{1}{t}\,dt$

54. $F(x) = \int_{x^3}^{0} t^3\,dt$

55. $F(x) = \frac{x}{x^2}(t+2)\,dt$

56. $F(x) = \int_{\ln x}^{e^x} \sin t\,dt$

5.5: Numerical Integration

Terms and Concepts

1. T/F: Simpson's Rule is a method of approximating antiderivatives.

2. What are the two basic situations where approximating the value of a definite integral is necessary?

3. Why are the Left and Right Hand Rules rarely used?

Problems

In Exercises 4-11, a definite integral is given.
(a) Approximating the definite integral with the Trapezoidal Rule and $n = 4$.
(b) Approximate the definite integral with Simpson's Rule and $n = 4$.
(c) Find the exact value of the integral.

4. $\int_{-1}^{1} x^2\,dx$

5. $\int_0^{10} 5x\,dx$

6. $\int_0^{\pi} \sin x\,dx$

7. $\int_0^4 \sqrt{x}\,dx$

8. $\int_0^3 (x^3 + 2x^2 - 5x + 7)\,dx$

9. $\int_0^1 x^4\,dx$

10. $\int_0^{2\pi} \cos x\,dx$

11. $\int_{-3}^{3} \sqrt{9 - x^2}\,dx$

In Exercises 12-19, approximate the definite integral with the Trapezoidal Rule and Simpson's Rule, with $n = 6$.

12. $\int_0^1 \cos(x^2)\,dx$

13. $\int_{-1}^{1} e^{x^2}\,dx$

14. $\int_0^5 \sqrt{x^2+1}\,dx$

15. $\int_0^{\pi} x \sin x\,dx$

16. $\int_0^{\pi/2} \sqrt{\cos x}\,dx$

17. $\int_1^4 \ln x\,dx$

18. $\int_{-1}^{1} \frac{1}{\sin x + 2}\,dx$

19. $\int_0^6 \frac{1}{\sin x + 2}\,dx$

In Exercises 20-23, find *n* such that the error in approximating the given definite integral is less than 0.0001 when using:
(a) the Trapezoidal Rule
(b) Simpson's Rule

20. $\int_0^\pi \sin x \, dx$

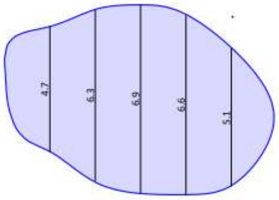
21. $\int_1^4 \frac{1}{\sqrt{x}} \, dx$

22. $\int_0^\pi \cos(x^2) \, dx$

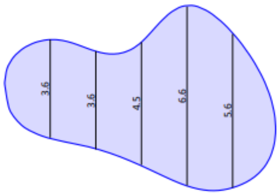
23. $\int_0^5 x^4 \, dx$

In Exercises 24-25, a region is given. Find the area of the region using Simpson's Rule:
(a) where the measurements are in centimeters, taken in 1cm increments, and
(b) where the measurements are in hundreds of yards, taken in 100 yd increments.

24.



25.



5.E: Applications of Integration (Exercises) is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

6: Techniques of Integration

6.1: Substitution
6.2: Integration by Parts
6.3: Trigonometric Integrals
6.4: Trigonometric Substitution
6.5: Partial Fraction Decomposition
6.6: Hyperbolic Functions
6.7: L'Hopital's Rule
6.8: Improper Integration
6.E: Applications of Antidifferentiation (Exercises)
Index

Thumbnail: Graph showing a few iterations of Newton's method on the graph $y = x^2$ *with initial guess of* $x_0 = 4$. *(Public Domain; Paul Breen).*

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 6: Techniques of Integration is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

6.1: Substitution

The previous chapter introduced the antiderivative and connected it to signed areas under a curve through the Fundamental Theorem of Calculus. The next chapter explores more applications of definite integrals than just area. As evaluating definite integrals will become important, we will want to find antiderivatives of a variety of functions.

This chapter is devoted to exploring techniques of antidifferentiation. While not every function has an antiderivative in terms of elementary functions (a concept introduced in the section on Numerical Integration), we can still find antiderivatives of a wide variety of functions.

Substitution

We motivate this section with an example. Let $f(x)=(x^2+3x-5)^{10}$. We can compute $f'(x)$ using the Chain Rule. It is:

$$f'(x)=10(x^2+3x-5)^9\cdot(2x+3)=(20x+30)(x^2+3x-5)^9. \tag{6.1.1}$$

Now consider this: What is $\int(20x+30)(x^2+3x-5)^9\ dx$? We have the answer in front of us;

$$\int(20x+30)(x^2+3x-5)^9\ dx=(x^2+3x-5)^{10}+C. \tag{6.1.2}$$

How would we have evaluated this indefinite integral without starting with $f(x)$ as we did?

This section explores *integration by substitution.* It allows us to "undo the Chain Rule." Substitution allows us to evaluate the above integral without knowing the original function first.

The underlying principle is to rewrite a "complicated" integral of the form $\int f(x)\ dx$ as a not--so--complicated integral $\int h(u)\ du$. We'll formally establish later how this is done. First, consider again our introductory indefinite integral, $\int(20x+30)(x^2+3x-5)^9\ dx$. Arguably the most "complicated" part of the integrand is $(x^2+3x-5)^9$. We wish to make this simpler; we do so through a substitution. Let $u=x^2+3x-5$. Thus

$$(x^2+3x-5)^9=u^9. \tag{6.1.3}$$

We have established u as a function of x, so now consider the differential of u:

$$du=(2x+3)dx. \tag{6.1.4}$$

Keep in mind that $(2x+3)$ and dx are multiplied; the dx is not "just sitting there."

Return to the original integral and do some substitutions through algebra:

$$\begin{aligned}\int(20x+30)(x^2+3x-5)^9\ dx&=\int 10(2x+3)(x^2+3x-5)^9\ dx && (6.1.5)\\ &=\int 10(\underbrace{x^2+3x-5}_{u})^9\ \underbrace{(2x+3)\ dx}_{du} && (6.1.6)\\ &=\int 10u^9\ du && (6.1.7)\\ &=u^{10}+C\quad(\text{replace } u \text{ with } x^2+3x-5) && (6.1.8)\\ &=(x^2+3x-5)^{10}+C && (6.1.9)\end{aligned}$$

One might well look at this and think "I (sort of) followed how that worked, but I could never come up with that on my own," but the process is learnable. This section contains numerous examples through which the reader will gain understanding and mathematical maturity enabling them to regard substitution as a natural tool when evaluating integrals.

We stated before that integration by substitution "undoes" the Chain Rule. Specifically, let $F(x)$ and $g(x)$ be differentiable functions and consider the derivative of their composition:

$$\frac{d}{dx}\Big(F\big(g(x)\big)\Big)=F'(g(x))g'(x). \tag{6.1.10}$$

Thus

$$\int F'(g(x))g'(x)\ dx=F(g(x))+C. \tag{6.1.11}$$

Integration by substitution works by recognizing the "inside" function $g(x)$ and replacing it with a variable. By setting $u = g(x)$, we can rewrite the derivative as

$$\frac{d}{dx}\Big(F\big(u\big)\Big) = F'(u)u'. \tag{6.1.12}$$

Since $du = g'(x)dx$, we can rewrite the above integral as

$$\int F'(g(x))g'(x)\ dx = \int F'(u)du = F(u) + C = F(g(x)) + C. \tag{6.1.13}$$

This concept is important so we restate it in the context of a theorem.

Theorem 6.1.1: Integration by Substitution

Let F and g be differentiable functions, where the range of g is an interval I contained in the domain of F. Then

$$\int F'(g(x))g'(x)\ dx = F(g(x)) + C. \tag{6.1.14}$$

If $u = g(x)$, then $du = g'(x)dx$ and

$$\int F'(g(x))g'(x)\ dx = \int F'(u)\ du = F(u) + C = F(g(x)) + C. \tag{6.1.15}$$

The point of substitution is to make the integration step easy. Indeed, the step $\int F'(u)\ du = F(u) + C$ looks easy, as the antiderivative of the derivative of F is just F, plus a constant. The "work" involved is making the proper substitution. There is not a step-by-step process that one can memorize; rather, experience will be one's guide. To gain experience, we now embark on many examples.

Example 6.1.1: Integrating by substitution

Evaluate $\int x\sin(x^2+5)\ dx$.

Solution

Knowing that substitution is related to the Chain Rule, we choose to let u be the "inside" function of $\sin(x^2+5)$. (This is not *always* a good choice, but it is often the best place to start.)

Let $u = x^2+5$, hence $du = 2x\ dx$. The integrand has an $x\ dx$ term, but not a $2x\ dx$ term. (Recall that multiplication is commutative, so the x does not physically have to be next to dx for there to be an $x\ dx$ term.) We can divide both sides of the du expression by 2:

$$du = 2x\,dx \quad \Rightarrow \quad \frac{1}{2}du = x\,dx. \tag{6.1.16}$$

We can now substitute.

$$\int x\sin(x^2+5)\ dx = \int \sin(\underbrace{x^2+5}_{u})\ \underbrace{x\ dx}_{\frac{1}{2}du} \tag{6.1.17}$$

$$= \int \frac{1}{2}\sin u\ du \tag{6.1.18}$$

$$= -\frac{1}{2}\cos u + C \quad \text{(now replace } u \text{ with } x^2+5\text{)} \tag{6.1.19}$$

$$= -\frac{1}{2}\cos(x^2+5) + C. \tag{6.1.20}$$

Thus $\int x\sin(x^2+5)\ dx = -\frac{1}{2}\cos(x^2+5) + C$. We can check our work by evaluating the derivative of the right hand side.

Example 6.1.2: Integrating by substitution

Evaluate $\int \cos(5x)\ dx$.

Solution

Again let u replace the "inside" function. Letting $u = 5x$, we have $du = 5dx$. Since our integrand does not have a $5dx$ term, we can divide the previous equation by 5 to obtain $\frac{1}{5}du = dx$. We can now substitute.

$$\int \cos(5x)\ dx = \int \cos(\underbrace{5x}_{u})\ \underbrace{dx}_{\frac{1}{5}du} \tag{6.1.21}$$

$$= \int \frac{1}{5}\cos u\ du \tag{6.1.22}$$

$$= \frac{1}{5}\sin u + C \tag{6.1.23}$$

$$= \frac{1}{5}\sin(5x) + C. \tag{6.1.24}$$

We can again check our work through differentiation.

The previous example exhibited a common, and simple, type of substitution. The "inside" function was a linear function (in this case, $y = 5x$). When the inside function is linear, the resulting integration is very predictable, outlined here.

Key Idea 10: Substitution With A Linear Function

Consider $\int F'(ax+b)\ dx$, where $a \neq 0$ and b are constants. Letting $u = ax+b$ gives $du = a \cdot dx$, leading to the result

$$\int F'(ax+b)\ dx = \frac{1}{a}F(ax+b) + C. \tag{6.1.25}$$

Thus $\int \sin(7x-4)\ dx = -\frac{1}{7}\cos(7x-4) + C$. Our next example can use Key Idea 10, but we will only employ it after going through all of the steps.

Example 6.1.3: Integrating by substituting a linear function

Evaluate $\int \frac{7}{-3x+1}\ dx$.

Solution

View this a composition of functions $f(g(x))$, where $f(x) = 7/x$ and $g(x) = -3x+1$. Employing our understanding of substitution, we let $u = -3x+1$, the inside function. Thus $du = -3dx$. The integrand lacks a -3; hence divide the previous equation by -3 to obtain $-du/3 = dx$. We can now evaluate the integral through substitution.

$$\int \frac{7}{-3x+1}\ dx = \int \frac{7}{u}\frac{du}{-3} \tag{6.1.26}$$

$$= \frac{-7}{3}\int \frac{du}{u} \tag{6.1.27}$$

$$= \frac{-7}{3}\ln|u| + C \tag{6.1.28}$$

$$= -\frac{7}{3}\ln|-3x+1| + C. \tag{6.1.29}$$

Using Key Idea 10 is faster, recognizing that u is linear and $a = -3$. One may want to continue writing out all the steps until they are comfortable with this particular shortcut.

Not all integrals that benefit from substitution have a clear "inside" function. Several of the following examples will demonstrate ways in which this occurs.

Example 6.1.4: Integrating by substitution

Evaluate $\int \sin x \cos x\ dx$.

Solution

There is not a composition of function here to exploit; rather, just a product of functions. Do not be afraid to experiment; when given an integral to evaluate, it is often beneficial to think "If I let u be *this*, then du must be *that* ..." and see if this helps simplify the integral at all.

In this example, let's set $u = \sin x$. Then $du = \cos x\ dx$, which we have as part of the integrand! The substitution becomes very straightforward:

$$\begin{aligned} \int \sin x \cos x\ dx &= \int u\ du && (6.1.30) \\ &= \frac{1}{2}u^2 + C && (6.1.31) \\ &= \frac{1}{2}\sin^2 x + C. && (6.1.32) \end{aligned}$$

One would do well to ask "What would happen if we let $u = \cos x$?" The result is just as easy to find, yet looks very different. The challenge to the reader is to evaluate the integral letting $u = \cos x$ and discover why the answer is the same, yet looks different.

Our examples so far have required "basic substitution." The next example demonstrates how substitutions can be made that often strike the new learner as being "nonstandard."

Example 6.1.5: Integrating by substitution

Evaluate $\int x\sqrt{x+3}\ dx$.

Solution

Recognizing the composition of functions, set $u = x + 3$. Then $du = dx$, giving what seems initially to be a simple substitution. But at this stage, we have:

$$\int x\sqrt{x+3}\ dx = \int x\sqrt{u}\ du. \tag{6.1.33}$$

We cannot evaluate an integral that has both an x and an u in it. We need to convert the x to an expression involving just u.

Since we set $u = x + 3$, we can also state that $u - 3 = x$. Thus we can replace x in the integrand with $u - 3$. It will also be helpful to rewrite $\sqrt{u}$ as $u^{\frac{1}{2}}$.

$$\begin{aligned} \int x\sqrt{x+3}\ dx &= \int (u-3)u^{\frac{1}{2}}\ du && (6.1.34) \\ &= \int \left(u^{\frac{3}{2}} - 3u^{\frac{1}{2}}\right)\ du && (6.1.35) \\ &= \frac{2}{5}u^{\frac{5}{2}} - 2u^{\frac{3}{2}} + C && (6.1.36) \\ &= \frac{2}{5}(x+3)^{\frac{5}{2}} - 2(x+3)^{\frac{3}{2}} + C. && (6.1.37) \end{aligned}$$

Checking your work is always a good idea. In this particular case, some algebra will be needed to make one's answer match the integrand in the original problem.

Example 6.1.6: Integrating by substitution

Evaluate $\int \frac{1}{x\ln x}\ dx$.

Solution

This is another example where there does not seem to be an obvious composition of functions. The line of thinking used in Example 6.1.5 is useful here: choose something for u and consider what this implies du must be. If u can be chosen such that du also appears in the integrand, then we have chosen well.

Choosing $u = 1/x$ makes $du = -1/x^2\ dx$; that does not seem helpful. However, setting $u = \ln x$ makes $du = 1/x\ dx$, which is part of the integrand. Thus:

$$\int \frac{1}{x\ln x}\ dx = \int \underbrace{\frac{1}{\ln x}}\ \underbrace{\frac{1}{x}\ dx} \tag{6.1.38}$$

$$1/u \qquad du$$

$$= \int \frac{1}{u}\, du \tag{6.1.39}$$

$$= \ln|u| + C \tag{6.1.40}$$

$$= \ln|\ln x| + C. \tag{6.1.41}$$

The final answer is interesting; the natural log of the natural log. Take the derivative to confirm this answer is indeed correct.

Integrals Involving Trigonometric Functions

Section 6.3 delves deeper into integrals of a variety of trigonometric functions; here we use substitution to establish a foundation that we will build upon.

The next three examples will help fill in some missing pieces of our antiderivative knowledge. We know the antiderivatives of the sine and cosine functions; what about the other standard functions tangent, cotangent, secant and cosecant? We discover these next.

Example $6.1.7$: Integration by substitution: antiderivatives of $\tan x$

Evaluate $\int \tan x\ dx$.

Solution

The previous paragraph established that we did not know the antiderivatives of tangent, hence we must assume that we have learned something in this section that can help us evaluate this indefinite integral.

Rewrite $\tan x$ as $\sin x/\cos x$. While the presence of a composition of functions may not be immediately obvious, recognize that $\cos x$ is "inside" the $1/x$ function. Therefore, we see if setting $u = \cos x$ returns usable results. We have that $du = -\sin x\ dx$, hence $-du = \sin x\ dx$. We can integrate:

$$\int \tan x\ dx = \int \frac{\sin x}{\cos x}\ dx \tag{6.1.42}$$

$$= \int \frac{1}{\underbrace{\cos x}_{u}} \underbrace{\sin x\ dx}_{-du} \tag{6.1.43}$$

$$= \int \frac{-1}{u}\ du \tag{6.1.44}$$

$$= -\ln|u| + C \tag{6.1.45}$$

$$= -\ln|\cos x| + C. \tag{6.1.46}$$

Some texts prefer to bring the -1 inside the logarithm as a power of $\cos x$, as in:

$$-\ln|\cos x| + C = \ln|(\cos x)^{-1}| + C \tag{6.1.47}$$

$$= \ln\left|\frac{1}{\cos x}\right| + C \tag{6.1.48}$$

$$= \ln|\sec x| + C. \tag{6.1.49}$$

Thus the result they give is $\int \tan x\ dx = \ln|\sec x| + C$. These two answers are equivalent.

Example $6.1.8$: Integrating by substitution: antiderivatives of $\sec x$

Evaluate $\int \sec x\ dx$.

Solution

This example employs a wonderful trick: multiply the integrand by "1" so that we see how to integrate more clearly. In this case, we write "1" as

$$1 = \frac{\sec x + \tan x}{\sec x + \tan x}. \tag{6.1.50}$$

This may seem like it came out of left field, but it works beautifully. Consider:

$$\int \sec x\ dx = \int \sec x \cdot \frac{\sec x + \tan x}{\sec x + \tan x}\ dx \tag{6.1.51}$$

$$= \int \frac{\sec^2 x + \sec x \tan x}{\sec x + \tan x}\ dx. \tag{6.1.52}$$

Now let $u=\sec x+\tan x$; this means $du=(\sec x\tan x+\sec^2 x)\ dx$, which is our numerator. Thus:

$$=\int \frac{du}{u} \tag{6.1.53}$$

$$=\ln|u|+C \tag{6.1.54}$$

$$=\ln|\sec x+\tan x|+C. \tag{6.1.55}$$

We can use similar techniques to those used in Examples 6.1.6 and 6.1.7 to find antiderivatives of $\cot x$ and $\csc x$ (which the reader can explore in the exercises.) We summarize our results here.

Theorem 6.1.1: Antiderivatives of Trigonometric Functions

1. $\int \sin x\ dx=-\cos x+C$
2. $\int \cos x\ dx=\sin x+C$
3. $\int \tan x\ dx\ =-\ln|\cos x|+C$
4. $\int \csc x\ dx=-\ln|\csc x+\cot x|+C$
5. $\int \sec x\ dx=\ln|\sec x+\tan x|+C$
6. $\int \cot x\ dx=\ln|\sin x|+C$

We explore one more common trigonometric integral.

Example 6.1.9: Integration by substitution: powers of $\cos x$ and $\sin x$

Evaluate $\int \cos^2 x\ dx$.

Solution

We have a composition of functions as $\cos^2 x=\left(\cos x\right)^2$.

However, setting $u=\cos x$ means $du=-\sin x\ dx$, which we do not have in the integral. Another technique is needed.

The process we'll employ is to use a Power Reducing formula for $\cos^2 x$ (perhaps consult the back of this text for this formula), which states

$$\cos^2 x=\frac{1+\cos(2x)}{2}. \tag{6.1.56}$$

The right hand side of this equation is not difficult to integrate. We have:

$$\int \cos^2 x\ dx=\int \frac{1+\cos(2x)}{2}\ dx \tag{6.1.57}$$

$$=\int \left(\frac{1}{2}+\frac{1}{2}\cos(2x)\right)\ dx. \tag{6.1.58}$$

Now use Key Idea 10:

$$=\frac{1}{2}x+\frac{1}{2}\frac{\sin(2x)}{2}+C \tag{6.1.59}$$

$$=\frac{1}{2}x+\frac{\sin(2x)}{4}+C. \tag{6.1.60}$$

We'll make significant use of this power--reducing technique in future sections.

Simplifying the Integrand

It is common to be reluctant to manipulate the integrand of an integral; at first, our grasp of integration is tenuous and one may think that working with the integrand will improperly change the results. Integration by substitution works using a different logic: as long as *equality* is maintained, the integrand can be manipulated so that its *form* is easier to deal with. The next two examples demonstrate common ways in which using algebra first makes the integration easier to perform.

Example 6.1.10: Integration by substitution: simplifying first

Evaluate $\displaystyle\int \frac{x^3+4x^2+8x+5}{x^2+2x+1}\,dx$.

Solution

One may try to start by setting u equal to either the numerator or denominator; in each instance, the result is not workable.

When dealing with rational functions (i.e., quotients made up of polynomial functions), it is an almost universal rule that everything works better when the degree of the numerator is less than the degree of the denominator. Hence we use polynomial division.

We skip the specifics of the steps, but note that when x^2+2x+1 is divided into x^3+4x^2+8x+5, it goes in $x+2$ times with a remainder of $3x+3$. Thus

$$\frac{x^3+4x^2+8x+5}{x^2+2x+1} = x+2+\frac{3x+3}{x^2+2x+1}. \tag{6.1.61}$$

Integrating $x+2$ is simple. The fraction can be integrated by setting $u=x^2+2x+1$, giving $du=(2x+2)\,dx$. This is very similar to the numerator. Note that $du/2=(x+1)\,dx$ and then consider the following:

$$\begin{aligned}
\int \frac{x^3+4x^2+8x+5}{x^2+2x+1}\,dx &= \int \left(x+2+\frac{3x+3}{x^2+2x+1}\right)\,dx && (6.1.62)\\
&= \int (x+2)\,dx + \int \frac{3(x+1)}{x^2+2x+1}\,dx && (6.1.63)\\
&= \frac{1}{2}x^2+2x+C_1+\int \frac{3}{u}\frac{du}{2} && (6.1.64)\\
&= \frac{1}{2}x^2+2x+C_1+\frac{3}{2}\ln|u|+C_2 && (6.1.65)\\
&= \frac{1}{2}x^2+2x+\frac{3}{2}\ln|x^2+2x+1|+C. && (6.1.66)
\end{aligned}$$

In some ways, we "lucked out" in that after dividing, substitution was able to be done. In later sections we'll develop techniques for handling rational functions where substitution is not directly feasible.

Example 6.1.11: Integration by alternate methods

Evaluate $\displaystyle\int \frac{x^2+2x+3}{\sqrt{x}}\,dx$ with, and without, substitution.

Solution

We already know how to integrate this particular example. Rewrite $\sqrt{x}$ as $x^{\frac{1}{2}}$ and simplify the fraction:

$$\frac{x^2+2x+3}{x^{1/2}} = x^{\frac{3}{2}}+2x^{\frac{1}{2}}+3x^{-\frac{1}{2}}. \tag{6.1.67}$$

We can now integrate using the Power Rule:

$$\begin{aligned}
\int \frac{x^2+2x+3}{x^{1/2}}\,dx &= \int \left(x^{\frac{3}{2}}+2x^{\frac{1}{2}}+3x^{-\frac{1}{2}}\right)\,dx && (6.1.68)\\
&= \frac{2}{5}x^{\frac{5}{2}}+\frac{4}{3}x^{\frac{3}{2}}+6x^{\frac{1}{2}}+C && (6.1.69)
\end{aligned}$$

This is a perfectly fine approach. We demonstrate how this can also be solved using substitution as its implementation is rather clever.

Let $u=\sqrt{x}=x^{\frac{1}{2}}$; therefore

$$du = \frac{1}{2}x^{-\frac{1}{2}}dx = \frac{1}{2\sqrt{x}}\,dx \quad\Rightarrow\quad 2du = \frac{1}{\sqrt{x}}\,dx. \tag{6.1.70}$$

This gives us $\int \frac{x^2+2x+3}{\sqrt{x}}\ dx = \int (x^2+2x+3)\cdot 2\ du$. What are we to do with the other x terms? Since $u = x^{\frac{1}{2}}$, $u^2 = x$, etc. We can then replace x^2 and x with appropriate powers of u. We thus have

$$\begin{aligned}\int \frac{x^2+2x+3}{\sqrt{x}}\ dx &= \int (x^2+2x+3)\cdot 2\ du \\ &= \int 2(u^4+2u^2+3)\ du \\ &= \frac{2}{5}u^5+\frac{4}{3}u^3+6u+C \\ &= \frac{2}{5}x^{\frac{5}{2}}+\frac{4}{3}x^{\frac{3}{2}}+6x^{\frac{1}{2}}+C,\end{aligned}$$

which is obviously the same answer we obtained before. In this situation, substitution is arguably more work than our other method. The fantastic thing is that it works. It demonstrates how flexible integration is.

Substitution and Inverse Trigonometric Functions

When studying derivatives of inverse functions, we learned that

$$\frac{d}{dx}\left(\tan^{-1}x\right) = \frac{1}{1+x^2}. \tag{6.1.71}$$

Applying the Chain Rule to this is not difficult; for instance,

$$\frac{d}{dx}\left(\tan^{-1}5x\right) = \frac{5}{1+25x^2}. \tag{6.1.72}$$

We now explore how Substitution can be used to "undo" certain derivatives that are the result of the Chain Rule applied to Inverse Trigonometric functions. We begin with an example.

Example $6.1.12$: Integrating by substitution: inverse trigonometric functions

Evaluate $\int \frac{1}{25+x^2}\ dx$.

Solution

The integrand looks similar to the derivative of the arctangent function. Note:

$$\begin{aligned}\frac{1}{25+x^2} &= \frac{1}{25(1+\frac{x^2}{25})} \qquad (6.1.73)\\ &= \frac{1}{25(1+\left(\frac{x}{5}\right)^2)} \qquad (6.1.74)\\ &= \frac{1}{25}\frac{1}{1+\left(\frac{x}{5}\right)^2}. \qquad (6.1.75)\end{aligned}$$

Thus

$$\int \frac{1}{25+x^2}\ dx = \frac{1}{25}\int \frac{1}{1+\left(\frac{x}{5}\right)^2}\ dx. \tag{6.1.76}$$

This can be integrated using Substitution. Set $u = x/5$, hence $du = dx/5$ or $dx = 5du$. Thus

$$\begin{aligned}\int \frac{1}{25+x^2}\ dx &= \frac{1}{25}\int \frac{1}{1+\left(\frac{x}{5}\right)^2}\ dx \qquad (6.1.77)\\ &= \frac{1}{5}\int \frac{1}{1+u^2}\ du \qquad (6.1.78)\\ &= \frac{1}{5}\tan^{-1}u+C \qquad (6.1.79)\\ &= \frac{1}{5}\tan^{-1}\left(\frac{x}{5}\right)+C \qquad (6.1.80)\end{aligned}$$

Example $6.1.12$ demonstrates a general technique that can be applied to other integrands that result in inverse trigonometric functions. The results are summarized here.

Theorem $6.1.2$: Integrals Involving Inverse Trigonomentric Functions

Let $a > 0$.

1. $\int \frac{1}{a^2+x^2}\,dx = \frac{1}{a}\tan^{-1}\left(\frac{x}{a}\right) + C$
2. $\int \frac{1}{\sqrt{a^2-x^2}}\,dx = \sin^{-1}\left(\frac{x}{a}\right) + C$
3. $\int \frac{1}{x\sqrt{x^2-a^2}}\,dx = \frac{1}{a}\sec^{-1}\left(\frac{|x|}{a}\right) + C$

Let's practice using Theorem $6.1.2$.

Example $6.1.13$: Integrating by substitution: inverse trigonometric functions

Evaluate the given indefinite integrals.

$$\int \frac{1}{9+x^2}\,dx, \quad \int \frac{1}{x\sqrt{x^2-\frac{1}{100}}}\,dx \quad \text{and} \quad \int \frac{1}{\sqrt{5-x^2}}\,dx. \tag{6.1.81}$$

Solution

Each can be answered using a straightforward application of Theorem $6.1.2$.

$\int \frac{1}{9+x^2}\,dx = \frac{1}{3}\tan^{-1}\frac{x}{3} + C$, as $a = 3$.

$\int \frac{1}{x\sqrt{x^2-\frac{1}{100}}}\,dx = 10\sec^{-1}10x + C$, as $a = \frac{1}{10}$.

$\int \frac{1}{\sqrt{5-x^2}} = \sin^{-1}\frac{x}{\sqrt{5}} + C$, as $a = \sqrt{5}$.

Most applications of Theorem $6.1.2$ are not as straightforward. The next examples show some common integrals that can still be approached with this theorem.

Example $6.1.14$: Integrating by substitution: completing the square

Evaluate $\int \frac{1}{x^2-4x+13}\,dx$.

Solution

Initially, this integral seems to have nothing in common with the integrals in Theorem $6.1.2$. As it lacks a square root, it almost certainly is not related to arcsine or arcsecant. It is, however, related to the arctangent function.

We see this by *completing the square* in the denominator. We give a brief reminder of the process here.

Start with a quadratic with a leading coefficient of 1. It will have the form of $x^2 + bx + c$. Take 1/2 of b, square it, and add/subtract it back into the expression. I.e.,

$$x^2+bx+c = \underbrace{x^2+bx+\frac{b^2}{4}}_{(x+b/2)^2} - \frac{b^2}{4} + c \tag{6.1.82}$$

$$= \left(x+\frac{b}{2}\right)^2 + c - \frac{b^2}{4} \tag{6.1.83}$$

In our example, we take half of -4 and square it, getting 4. We add/subtract it into the denominator as follows:

$$\frac{1}{x^2-4x+13} = \frac{1}{\underbrace{x^2-4x+4}_{(x-2)^2}-4+13} \tag{6.1.84}$$

$$= \frac{1}{(x-2)^2+9} \tag{6.1.85}$$

We can now integrate this using the arctangent rule. Technically, we need to substitute first with $u = x - 2$, but we can employ Key Idea 10 instead. Thus we have

$$\int \frac{1}{x^2 - 4x + 13}\ dx = \int \frac{1}{(x-2)^2 + 9}\ dx = \frac{1}{3}\tan^{-1}\frac{x-2}{3} + C. \tag{6.1.86}$$

Example 6.1.15: Integrals requiring multiple methods

Evaluate $\int \frac{4-x}{\sqrt{16-x^2}}\ dx$.

Solution

This integral requires two different methods to evaluate it. We get to those methods by splitting up the integral:

$$\int \frac{4-x}{\sqrt{16-x^2}}\ dx = \int \frac{4}{\sqrt{16-x^2}}\ dx - \int \frac{x}{\sqrt{16-x^2}}\ dx. \tag{6.1.87}$$

The first integral is handled using a straightforward application of Theorem 6.1.2; the second integral is handled by substitution, with $u = 16 - x^2$. We handle each separately.

$\int \frac{4}{\sqrt{16-x^2}}\ dx = 4\sin^{-1}\frac{x}{4} + C.$

$\int \frac{x}{\sqrt{16-x^2}}\ dx$: Set $u = 16 - x^2$, so $du = -2x dx$ and $x dx = -du/2$. We have

$$\begin{aligned}\int \frac{x}{\sqrt{16-x^2}}\ dx &= \int \frac{-du/2}{\sqrt{u}} &(6.1.88)\\ &= -\frac{1}{2}\int \frac{1}{\sqrt{u}}\ du &(6.1.89)\\ &= -\sqrt{u} + C &(6.1.90)\\ &= -\sqrt{16-x^2} + C. &(6.1.91)\end{aligned}$$

Combining these together, we have

$$\int \frac{4-x}{\sqrt{16-x^2}}\ dx = 4\sin^{-1}\frac{x}{4} + \sqrt{16-x^2} + C. \tag{6.1.92}$$

Substitution and Definite Integration

This section has focused on evaluating indefinite integrals as we are learning a new technique for finding antiderivatives. However, much of the time integration is used in the context of a definite integral. Definite integrals that require substitution can be calculated using the following workflow:

1. Start with a definite integral $\int_a^b f(x)\ dx$ that requires substitution.
2. Ignore the bounds; use substitution to evaluate $\int f(x)\ dx$ and find an antiderivative $F(x)$.
3. Evaluate $F(x)$ at the bounds; that is, evaluate $F(x)\Big|_a^b = F(b) - F(a)$.

This workflow works fine, but substitution offers an alternative that is powerful and amazing (and a little time saving).

At its heart, (using the notation of Theorem 6.1.1 substitution converts integrals of the form $\int F'(g(x))g'(x)\ dx$ into an integral of the form $\int F'(u)\ du$ with the substitution of $u = g(x)$. The following theorem states how the bounds of a definite integral can be changed as the substitution is performed.

Theorem 6.1.3: Substitution with Definite Integrals

Let F and g be differentiable functions, where the range of g is an interval I that is contained in the domain of F. Then

$$\int_a^b F'\big(g(x)\big)g'(x)\ dx = \int_{g(a)}^{g(b)} F'(u)\ du. \tag{6.1.93}$$

In effect, Theorem 6.1.3 states that once you convert to integrating with respect to u, you do not need to switch back to evaluating with respect to x. A few examples will help one understand.

Example 6.1.16: Definite integrals and substitution: changing the bounds

Evaluate $\int_0^2 \cos(3x-1)\ dx$ using Theorem 6.1.3.

Solution

Observing the composition of functions, let $u = 3x-1$, hence $du = 3dx.As\backslash(3dx\backslash)$ does not appear in the integrand, divide the latter equation by 3 to get $du/3 = dx$.

By setting $u = 3x-1$, we are implicitly stating that $g(x) = 3x-1$. Theorem 6.1.3 states that the new lower bound is $g(0) = -1$; the new upper bound is $g(2) = 5$. We now evaluate the definite integral:

$$\int_1^2 \cos(3x-1)\ dx = \int_{-1}^5 \cos u \frac{du}{3} \tag{6.1.94}$$

$$= \frac{1}{3}\sin u\Big|_{-1}^5 \tag{6.1.95}$$

$$= \frac{1}{3}\Big(\sin 5 - \sin(-1)\Big) \approx -0.039. \tag{6.1.96}$$

Notice how once we converted the integral to be in terms of u, we never went back to using x.

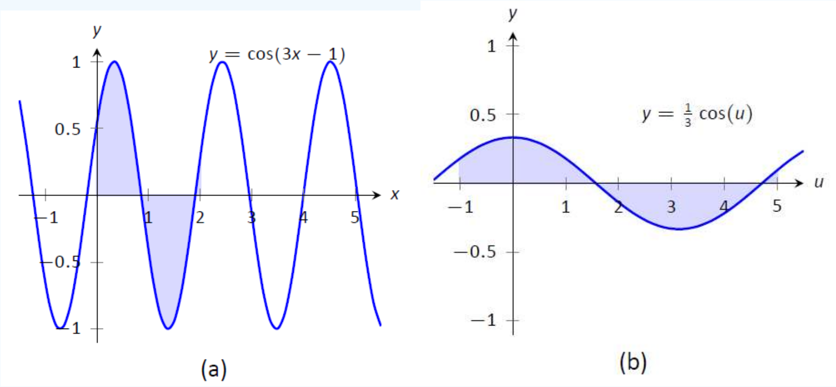


***Figure** 6.1.1: Graphing the areas defined by the definite integrals of Example 6.1.16*

The graphs in Figure 6.1.1 tell more of the story. In (a) the area defined by the original integrand is shaded, whereas in (b) the area defined by the new integrand is shaded. In this particular situation, the areas look very similar; the new region is "shorter" but "wider," giving the same area.

Example 6.1.17: Definite integrals and substitution: changing the bounds

Evaluate $\int_0^{\pi/2} \sin x \cos x\ dx$ using Theorem 6.1.3.

Solution

We saw the corresponding indefinite integral in Example 6.1.4. In that example we set $u = \sin x$ but stated that we could have let $u = \cos x$. For variety, we do the latter here.

Let $u = g(x) = \cos x$, giving $du = -\sin x\ dx$ and hence $\sin x\ dx = -du$. The new upper bound is $g(\pi/2) = 0$; the new lower bound is $g(0) = 1$. Note how the lower bound is actually larger than the upper bound now. We have

$$\int_0^{\pi/2} \sin x \cos x\ dx = \int_1^0 -u\ du \quad \text{(switch bounds \& change sign)} \tag{6.1.97}$$

$$= \int_0^1 u\ du \tag{6.1.98}$$

$$= \frac{1}{2}u^2\Big|_0^1 = 1/2. \tag{6.1.99}$$

In Figure 6.1.2 we have again graphed the two regions defined by our definite integrals. Unlike the previous example, they bear no resemblance to each other. However, Theorem 6.1.3 guarantees that they have the same area.

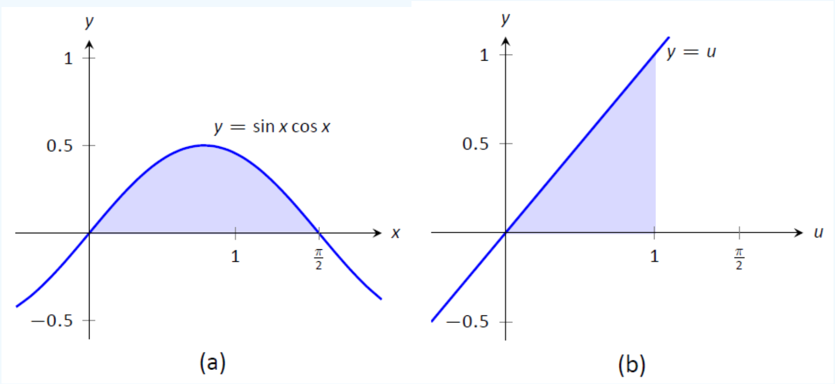


***Figure** 6.1.2: Graphing the areas defined by the definite integrals of Example 6.1.17.*

Integration by substitution is a powerful and useful integration technique. The next section introduces another technique, called Integration by Parts. As substitution "undoes" the Chain Rule, integration by parts "undoes" the Product Rule. Together, these two techniques provide a strong foundation on which most other integration techniques are based.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/
- Integrated by Justin Marshall.

This page titled 6.1: Substitution is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

6.2: Integration by Parts

Here's a simple integral that we can't yet evaluate:

$$\int x\cos x\,dx. \tag{6.2.1}$$

It's a simple matter to take the derivative of the integrand using the Product Rule, but there is no Product Rule for integrals. However, this section introduces *Integration by Parts*, a method of integration that is based on the Product Rule for derivatives. It will enable us to evaluate this integral.

The Product Rule says that if u and v are functions of x, then $(uv)' = u'v + uv'$. For simplicity, we've written u for $u(x)$ and v for $v(x)$. Suppose we integrate both sides with respect to x. This gives

$$\int (uv)'\,dx = \int (u'v + uv')\,dx. \tag{6.2.2}$$

By the Fundamental Theorem of Calculus, the left side integrates to uv. The right side can be broken up into two integrals, and we have

$$uv = \int u'v\,dx + \int uv'\,dx. \tag{6.2.3}$$

Solving for the second integral we have

$$\int uv'\,dx = uv - \int u'v\,dx. \tag{6.2.4}$$

Using differential notation, we can write $du = u'(x)dx$ and $dv = v'(x)dx$ and the expression above can be written as follows:

$$\int u\,dv = uv - \int v\,du. \tag{6.2.5}$$

This is the Integration by Parts formula. For reference purposes, we state this in a theorem.

Theorem $6.2.1$: Integration by Parts

Let u and v be differentiable functions of x on an interval I containing a and b. Then

$$\int u\ dv = uv - \int v\ du, \tag{6.2.6}$$

and *integration by parts*

$$\int_{x=a}^{x=b} u\ dv = uv\Big|_a^b - \int_{x=a}^{x=b} v\ du. \tag{6.2.7}$$

Let's try an example to understand our new technique.

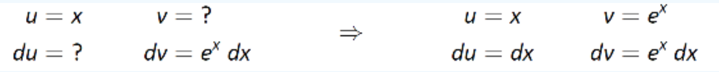
Example $6.2.1$: Integrating using Integration by Parts

Evaluate $\displaystyle\int x\cos x\ dx$.

Solution

The key to Integration by Parts is to identify part of the integrand as "u" and part as "dv." Regular practice will help one make good identifications, and later we will introduce some principles that help. For now, let $u = x$ and $dv = \cos x\ dx$.

It is generally useful to make a small table of these values as done below. Right now we only know u and dv as shown on the left of Figure $6.2.1$; on the right we fill in the rest of what we need. If $u = x$, then $du = dx$. Since $dv = \cos x\ dx$, v is an antiderivative of $\cos x$. We choose $v = \sin x$.



$$\begin{array}{ll} u = x & v = ? \\ du = ? & dv = e^x\ dx \end{array} \quad \Rightarrow \quad \begin{array}{ll} u = x & v = e^x \\ du = dx & dv = e^x\ dx \end{array}$$

***Figure** 6.2.1: Setting up integration by parts.*

Now substitute all of this into the Integration by Parts formula, giving

$$\int x\cos x\,dx = x\sin x - \int \sin x\,dx. \tag{6.2.8}$$

We can then integrate $\sin x$ to get $-\cos x + C$ and overall our answer is

$$\int x\cos x\;dx = x\sin x + \cos x + C. \tag{6.2.9}$$

Note how the antiderivative contains a product, $x\sin x$. This product is what makes Integration by Parts necessary.

The example above demonstrates how Integration by Parts works in general. We try to identify u and dv in the integral we are given, and the key is that we usually want to choose u and dv so that du is simpler than u and v is hopefully not too much more complicated than dv. This will mean that the integral on the right side of the Integration by Parts formula, $\int v\,du$ will be simpler to integrate than the original integral $\int u\,dv$.

In the example above, we chose $u = x$ and $dv = \cos x\,dx$. Then $du = dx$ was simpler than u and $v = \sin x$ is no more complicated than dv. Therefore, instead of integrating $x\cos x\,dx$, we could integrate $\sin x\,dx$, which we knew how to do.

A useful mnemonic for helping to determine u is "LIATE," where

$$L = \mathbf{L}ogarithmic, I = \mathbf{I}nverseTrig., A = \mathbf{A}lgebraic(polynomials), \tag{6.2.10}$$

$$T = \mathbf{T}rigonometric, and E = \mathbf{E}xponential. \tag{6.2.11}$$

If the integrand contains both a logarithmic and an algebraic term, in general letting u be the logarithmic term works best, as indicated by L coming before A in LIATE.

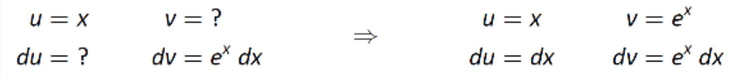
We now consider another example.

Example 6.2.2: Integrating using Integration by Parts

Evaluate $\int xe^x\,dx$.

Solution

The integrand contains an **A**lgebraic term (x) and an \textbf{E}xponential term (e^x). Our mnemonic suggests letting u be the algebraic term, so we choose $u = x$ and $dv = e^x\,dx$. Then $du = dx$ and $v = e^x$ as indicated by the tables below.



$u = x$	$v = ?$	$\Rightarrow$	$u = x$	$v = e^x$
$du = ?$	$dv = e^x\,dx$		$du = dx$	$dv = e^x\,dx$

***Figure** 6.2.2: Setting up Integration by Parts.*

We see du is simpler than u, while there is no change in going from dv to v. This is good. The Integration by Parts formula gives

$$\int xe^x\,dx = xe^x - \int e^x\,dx. \tag{6.2.12}$$

The integral on the right is simple; our final answer is

$$\int xe^x\;dx = xe^x - e^x + C. \tag{6.2.13}$$

Note again how the antiderivatives contain a product term.

Example 6.2.3: Integrating using Integration by Parts

Evaluate $\int x^2\cos x\,dx$.

Solution

The mnemonic suggests letting $u = x^2$ instead of the trigonometric function, hence $dv = \cos x\,dx$. Then $du = 2x\,dx$ and $v = \sin x$ as shown below.

$$\begin{array}{ll} u = x^2 & v = ? \\ du = ? & dv = \cos x\,dx \end{array} \quad \Rightarrow \quad \begin{array}{ll} u = x^2 & v = \sin x \\ du = 2x\,dx & dv = \cos x\,dx \end{array}$$

Figure *6.2.3**:** Setting up Integration by Parts.*

The Integration by Parts formula gives

$$\int x^2 \cos x\,dx = x^2 \sin x - \int 2x \sin x\,dx. \tag{6.2.14}$$

At this point, the integral on the right is indeed simpler than the one we started with, but to evaluate it, we need to do Integration by Parts again. Here we choose $u = 2x$ and $dv = \sin x$ and fill in the rest below.

$$\begin{array}{ll} u = 2x & v = ? \\ du = ? & dv = \sin x\,dx \end{array} \quad \Rightarrow \quad \begin{array}{ll} u = 2x & v = -\cos x \\ du = 2\,dx & dv = \sin x\,dx \end{array}$$

Figure *6.2.4**:** Setting up Integration by Parts.*

The integral all the way on the right is now something we can evaluate. It evaluates to $-2\sin x$. Then going through and simplifying, being careful to keep all the signs straight, our answer is

$$\int x^2 \cos x\ dx = x^2 \sin x + 2x\cos x - 2\sin x + C. \tag{6.2.15}$$

Example 6.2.4: Integrating using Integration by Parts

Evaluate $\int e^x \cos x\,dx$.

Solution

This is a classic problem. Our mnemonic suggests letting u be the trigonometric function instead of the exponential. In this particular example, one can let u be either $\cos x$ or e^x; to demonstrate that we do not have to follow LIATE, we choose $u = e^x$ and hence $dv = \cos x\,dx$. Then $du = e^x\,dx$ and $v = \sin x$ as shown below.

$$\begin{array}{ll} u = e^x & v = ? \\ du = ? & dv = \cos x\,dx \end{array} \quad \Rightarrow \quad \begin{array}{ll} u = e^x & v = \sin x \\ du = e^x\,dx & dv = \cos x\,dx \end{array}$$

Figure *6.2.5**:** Setting up Integration by Parts.*

Notice that du is no simpler than u, going against our general rule (but bear with us). The Integration by Parts formula yields

$$\int e^x \cos x\ dx = e^x \sin x - \int e^x \sin x\,dx. \tag{6.2.16}$$

The integral on the right is not much different than the one we started with, so it seems like we have gotten nowhere. Let's keep working and apply Integration by Parts to the new integral, using $u = e^x$ and $dv = \sin x\,dx$. This leads us to the following:

$$\begin{array}{ll} u = e^x & v = ? \\ du = ? & dv = \sin x\,dx \end{array} \quad \Rightarrow \quad \begin{array}{ll} u = e^x & v = -\cos x \\ du = e^x\,dx & dv = \sin x\,dx \end{array}$$

Figure *6.2.6: Setting up Integration by Parts.*

The Integration by Parts formula then gives:

$$\begin{aligned} \int e^x \cos x\,dx &= e^x \sin x - \left(-e^x \cos x - \int -e^x \cos x\,dx\right) \\ &= e^x \sin x + e^x \cos x - \int e^x \cos x\ dx. \end{aligned}$$

It seems we are back right where we started, as the right hand side contains $\int e^x \cos x\,dx$. But this is actually a good thing.

Add $\int e^x \cos x \; dx$ to both sides. This gives

$$2\int e^x \cos x \; dx = e^x \sin x + e^x \cos x$$
$$\text{Now divide both sides by 2:} \int e^x \cos x \; dx = \frac{1}{2}\big(e^x \sin x + e^x \cos x\big).$$

Simplifying a little and adding the constant of integration, our answer is thus

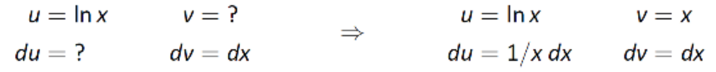
$$\int e^x \cos x \; dx = \frac{1}{2} e^x \left(\sin x + \cos x\right) + C. \tag{6.2.17}$$

Example $6.2.5$: Integrating using Integration by Parts: antiderivative of $\ln x$

Evaluate $\displaystyle\int \ln x \, dx$.

Solution

One may have noticed that we have rules for integrating the familiar trigonometric functions and e^x, but we have not yet given a rule for integrating $\ln x$. That is because $\ln x$ can't easily be integrated with any of the rules we have learned up to this point. But we can find its antiderivative by a clever application of Integration by Parts. Set $u = \ln x$ and $dv = dx$. This is a good, sneaky trick to learn as it can help in other situations. This determines $du = (1/x) \; dx$ and $v = x$ as shown below.



$$\begin{array}{ll} u = \ln x & v = ? \\ du = ? & dv = dx \end{array} \quad \Rightarrow \quad \begin{array}{ll} u = \ln x & v = x \\ du = 1/x \, dx & dv = dx \end{array}$$

Figure $6.2.7$***:*** *Setting up Integration by Parts.*

Putting this all together in the Integration by Parts formula, things work out very nicely:

$$\int \ln x \, dx = x \ln x - \int x \frac{1}{x} \, dx. \tag{6.2.18}$$

The new integral simplifies to $\int 1 \, dx$, which is about as simple as things get. Its integral is $x + C$ and our answer is

$$\int \ln x \; dx = x \ln x - x + C. \tag{6.2.19}$$

Example $6.2.6$: Integrating using Int. by Parts: antiderivative of $\arctan x$

Evaluate $\displaystyle\int \arctan x \, dx$.

Solution

The same sneaky trick we used above works here. Let $u = \arctan x$ and $dv = dx$. Then $du = 1/(1+x^2) \; dx$ and $v = x$. The Integration by Parts formula gives

$$\int \arctan x \, dx = x \arctan x - \int \frac{x}{1+x^2} \, dx. \tag{6.2.20}$$

The integral on the right can be solved by substitution. Taking $u = 1 + x^2$, we get $du = 2x \, dx$. The integral then becomes

$$\int \arctan x \, dx = x \arctan x - \frac{1}{2} \int \frac{1}{u} \, du. \tag{6.2.21}$$

The integral on the right evaluates to $\ln |u| + C$, which becomes $\ln(1 + x^2) + C$. Therefore, the answer is

$$\int \arctan x \; dx = x \arctan x - \ln(1 + x^2) + C. \tag{6.2.22}$$

Substitution Before Integration

When taking derivatives, it was common to employ multiple rules (such as using both the Quotient and the Chain Rules). It should then come as no surprise that some integrals are best evaluated by combining integration techniques. In particular, here we illustrate making an "unusual" substitution first before using Integration by Parts.

Example $6.2.7$: Integration by Parts after substitution

Evaluate $\int \cos(\ln x)\ dx$.

Solution

The integrand contains a composition of functions, leading us to think Substitution would be beneficial. Letting $u = \ln x$, we have $du = 1/x\ dx$. This seems problematic, as we do not have a $1/x$ in the integrand. But consider:

$$du = \frac{1}{x}\ dx \Rightarrow x \cdot du = dx. \tag{6.2.23}$$

Since $u = \ln x$, we can use inverse functions and conclude that $x = e^u$. Therefore we have that

$$\begin{aligned} dx &= x \cdot du \\ &= e^u\ du. \end{aligned}$$

We can thus replace $\ln x$ with u and dx with $e^u\ du$. Thus we rewrite our integral as

$$\int \cos(\ln x)\ dx = \int e^u \cos u\ du. \tag{6.2.24}$$

We evaluated this integral in Example $6.2.4$. Using the result there, we have:

$$\begin{aligned} \int \cos(\ln x)\ dx &= \int e^u \cos u\ du \\ &= \frac{1}{2} e^u \big(\sin u + \cos u \big) + C \\ &= \frac{1}{2} e^{\ln x} \big(\sin(\ln x) + \cos(\ln x) \big) + C \\ &= \frac{1}{2} x \big(\sin(\ln x) + \cos(\ln x) \big) + C. \end{aligned}$$

Definite Integrals and Integration By Parts

So far we have focused only on evaluating indefinite integrals. Of course, we can use Integration by Parts to evaluate definite integrals as well, as Theorem $6.2.1$ states. We do so in the next example.

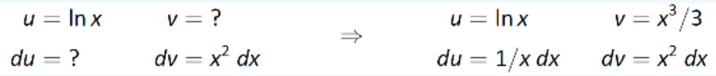
Example $6.2.8$: Definite integration using Integration by Parts

Evaluate $\int_1^2 x^2 \ln x\, dx$.

Solution

Our mnemonic suggests letting $u = \ln x$, hence $dv = x^2\, dx$.

We then get $du = (1/x)\, dx$ and $v = x^3/3$ as shown below.



$u = \ln x$	$v = ?$	$\Rightarrow$	$u = \ln x$	$v = x^3/3$
$du = ?$	$dv = x^2\ dx$		$du = 1/x\ dx$	$dv = x^2\ dx$

Figure $6.2.8$***:*** *Setting up Integration by Parts.*

The Integration by Parts formula then gives

$$
\begin{aligned}
\int_1^2 x^2 \ln x \, dx &= \frac{x^3}{3}\ln x\Big|_1^2 - \int_1^2 \frac{x^3}{3}\frac{1}{x}\,dx \\
&= \frac{x^3}{3}\ln x\Big|_1^2 - \int_1^2 \frac{x^2}{3}\,dx \\
&= \frac{x^3}{3}\ln x\Big|_1^2 - \frac{x^3}{9}\Big|_1^2 \\
&= \left(\frac{x^3}{3}\ln x - \frac{x^3}{9}\right)\Big|_1^2 \\
&= \left(\frac{8}{3}\ln 2 - \frac{8}{9}\right) - \left(\frac{1}{3}\ln 1 - \frac{1}{9}\right) \\
&= \frac{8}{3}\ln 2 - \frac{7}{9} \\
&\approx 1.07.
\end{aligned}
$$

In general, Integration by Parts is useful for integrating certain products of functions, like $\int xe^x\,dx$ or $\int x^3 \sin x\,dx$. It is also useful for integrals involving logarithms and inverse trigonometric functions.

As stated before, integration is generally more difficult than derivation. We are developing tools for handling a large array of integrals, and experience will tell us when one tool is preferable/necessary over another. For instance, consider the three similar--looking integrals

$$\int xe^x\,dx, \qquad \int xe^{x^2}\,dx \qquad \text{and} \qquad \int xe^{x^3}\,dx. \tag{6.2.25}$$

While the first is calculated easily with Integration by Parts, the second is best approached with Substitution. Taking things one step further, the third integral has no answer in terms of elementary functions, so none of the methods we learn in calculus will get us the exact answer.

Integration by Parts is a very useful method, second only to substitution. In the following sections of this chapter, we continue to learn other integration techniques. The next section focuses on handling integrals containing trigonometric functions.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/
- Integrated by Justin Marshall.

This page titled 6.2: Integration by Parts is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

6.3: Trigonometric Integrals

Functions involving trigonometric functions are useful as they are good at describing periodic behavior. This section describes several techniques for finding antiderivatives of certain combinations of trigonometric functions.

Integrals of the form $\int \sin^m x \cos^n x \ dx$

In learning the technique of Substitution, we saw the integral $\int \sin x \cos x \ dx$ in Example 6.1.4. The integration was not difficult, and one could easily evaluate the indefinite integral by letting $u = \sin x$ or by letting $u = \cos x$. This integral is easy since the power of both sine and cosine is 1.

We generalize this integral and consider integrals of the form $\int \sin^m x \cos^n x \ dx$, where m, n are nonnegative integers. Our strategy for evaluating these integrals is to use the identity $\cos^2 x + \sin^2 x = 1$ to convert high powers of one trigonometric function into the other, leaving a single sine or cosine term in the integrand. We summarize the general technique in the following Key Idea.

Key Idea 11: Integrals Involving Powers of Sine and Cosine

Consider $\int \sin^m x \cos^n x \ dx$, where m, n are nonnegative integers.

1. If m is odd, then $m = 2k+1$ for some integer k. Rewrite

$$\sin^m x = \sin^{2k+1} x = \sin^{2k} x \sin x = (\sin^2 x)^k \sin x = (1-\cos^2 x)^k \sin x. \tag{6.3.1}$$

Then

$$\int \sin^m x \cos^n x \ dx = \int (1-\cos^2 x)^k \sin x \cos^n x \ dx = -\int (1-u^2)^k u^n \ du, \tag{6.3.2}$$

where $u = \cos x$ and $du = -\sin x \ dx$.

2. If n is odd, then using substitutions similar to that outlined above we have

$$\int \sin^m x \cos^n x \ dx = \int u^m (1-u^2)^k \ du, \tag{6.3.3}$$

where $u = \sin x$ and $du = \cos x \ dx$.

3. If both m and n are even, use the power--reducing identities

$$\cos^2 x = \frac{1+\cos(2x)}{2} \quad \text{and} \quad \sin^2 x = \frac{1-\cos(2x)}{2} \tag{6.3.4}$$

to reduce the degree of the integrand. Expand the result and apply the principles of this Key Idea again.

We practice applying Key Idea 11 in the next examples.

Example 6.3.1: Integrating powers of sine and cosine

Evaluate $\int \sin^5 x \cos^8 x \ dx$.

Solution

The power of the sine term is odd, so we rewrite $\sin^5 x$ as

$$\sin^5 x = \sin^4 x \sin x = (\sin^2 x)^2 \sin x = (1-\cos^2 x)^2 \sin x.$$

Our integral is now $\int (1-\cos^2 x)^2 \cos^8 x \sin x \ dx$. Let $u = \cos x$, hence $du = -\sin x \ dx$. Making the substitution and expanding the integrand gives

$$\begin{aligned} \int (1-\cos^2)^2 \cos^8 x \sin x \ dx &= -\int (1-u^2)^2 u^8 \ du = -\int \left(1-2u^2+u^4\right) u^8 \ du \\ &= -\int \left(u^8 - 2u^{10} + u^{12}\right) du. \end{aligned}$$

This final integral is not difficult to evaluate, giving

$$\begin{aligned}-\int\left(u^8-2u^{10}+u^{12}\right)\,du &= -\frac{1}{9}u^9+\frac{2}{11}u^{11}-\frac{1}{13}u^{13}+C\\ &= -\frac{1}{9}\cos^9 x+\frac{2}{11}\cos^{11} x-\frac{1}{13}\cos^{13} x+C.\end{aligned}$$

Example 6.3.2: Integrating powers of sine and cosine

Evaluate $\int \sin^5 x\cos^9 x\ dx$.

Solution

The powers of both the sine and cosine terms are odd, therefore we can apply the techniques of Key Idea 11 to either power. We choose to work with the power of the cosine term since the previous example used the sine term's power.

We rewrite $\cos^9 x$ as

$$\begin{aligned}\cos^9 x &= \cos^8 x\cos x\\ &= (\cos^2 x)^4\cos x\\ &= (1-\sin^2 x)^4\cos x\\ &= (1-4\sin^2 x+6\sin^4 x-4\sin^6 x+\sin^8 x)\cos x.\end{aligned}$$

We rewrite the integral as

$$\int \sin^5 x\cos^9 x\ dx=\int \sin^5 x\left(1-4\sin^2 x+6\sin^4 x-4\sin^6 x+\sin^8 x\right)\cos x\ dx.$$

Now substitute and integrate, using $u=\sin x$ and $du=\cos x\ dx$.

$$\begin{aligned}\int u^5(1-4u^2+6u^4-4u^6+u^8)\ du &= \int\left(u^5-4u^7+6u^9-4u^{11}+u^{13}\right)\,du\\ &= \frac{1}{6}u^6-\frac{1}{2}u^8+\frac{3}{5}u^{10}-\frac{1}{3}u^{12}+\frac{1}{14}u^{14}+C\\ &= \frac{1}{6}\sin^6 x-\frac{1}{2}\sin^8 x+\frac{3}{5}\sin^{10} x+\ldots\\ &\quad -\frac{1}{3}\sin^{12} x+\frac{1}{14}\sin^{14} x+C.\end{aligned}$$

Technology Note: The work we are doing here can be a bit tedious, but the skills developed (problem solving, algebraic manipulation, etc.) are important. Nowadays problems of this sort are often solved using a computer algebra system. The powerful program Mathematica integrates $\int \sin^5 x\cos^9 x\ dx$ as

$$\begin{aligned}f(x)=&-\frac{45\cos(2x)}{16384}-\frac{5\cos(4x)}{8192}+\frac{19\cos(6x)}{49152}+\frac{\cos(8x)}{4096}\\ &-\frac{\cos(10x)}{81920}-\frac{\cos(12x)}{24576}-\frac{\cos(14x)}{114688},\end{aligned}\tag{6.3.5}$$

which clearly has a different form than our answer in Example 6.3.2, which is

$$g(x)=\frac{1}{6}\sin^6 x-\frac{1}{2}\sin^8 x+\frac{3}{5}\sin^{10} x-\frac{1}{3}\sin^{12} x+\frac{1}{14}\sin^{14} x.\tag{6.3.6}$$

Figure 6.3.1 shows a graph of f and g; they are clearly not equal, but they differ *only by a constant*. That is $g(x)=f(x)+C$ for some constant C. So we have two different antiderivatives of the same function, meaning both answers are correct.

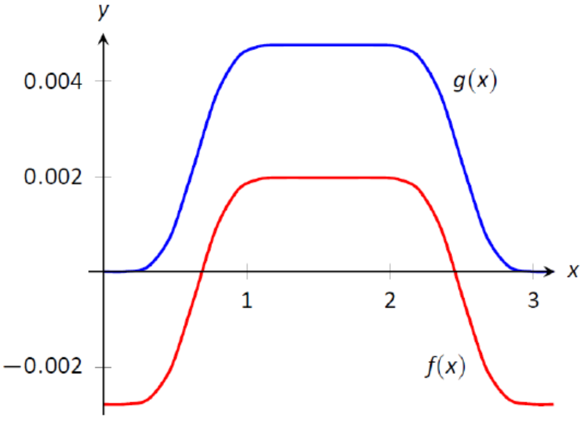


***Figure** $6.3.1$**:** A plot of $f(x)$ and $g(x)$ from Example $6.3.2$ and the Technology Note.*

Example $6.3.3$: Integrating powers of sine and cosine

Evaluate $\int \cos^4 x \sin^2 x\ dx$.

Solution

The powers of sine and cosine are both even, so we employ the power--reducing formulas and algebra as follows.

$$\begin{aligned}\int \cos^4 x \sin^2 x\ dx &= \int \left(\frac{1+\cos(2x)}{2}\right)^2 \left(\frac{1-\cos(2x)}{2}\right)\ dx \\ &= \int \frac{1+2\cos(2x)+\cos^2(2x)}{4} \cdot \frac{1-\cos(2x)}{2}\ dx \\ &= \int \frac{1}{8}\big(1+\cos(2x)-\cos^2(2x)-\cos^3(2x)\big)\ dx\end{aligned}$$

The $\cos(2x)$ term is easy to integrate, especially with Key Idea 10. The $\cos^2(2x)$ term is another trigonometric integral with an even power, requiring the power--reducing formula again. The $\cos^3(2x)$ term is a cosine function with an odd power, requiring a substitution as done before. We integrate each in turn below.

$$\int \cos(2x)\ dx = \frac{1}{2}\sin(2x) + C.$$

$$\int \cos^2(2x)\ dx = \int \frac{1+\cos(4x)}{2}\ dx = \frac{1}{2}\Big(x + \frac{1}{4}\sin(4x)\Big) + C.$$

Finally, we rewrite $\cos^3(2x)$ as

$$\cos^3(2x) = \cos^2(2x)\cos(2x) = \big(1-\sin^2(2x)\big)\cos(2x). \tag{6.3.7}$$

Letting $u = \sin(2x)$, we have $du = 2\cos(2x)\ dx$, hence

$$\begin{aligned}\int \cos^3(2x)\ dx &= \int \big(1-\sin^2(2x)\big)\cos(2x)\ dx \\ &= \int \frac{1}{2}(1-u^2)\ du \\ &= \frac{1}{2}\Big(u - \frac{1}{3}u^3\Big) + C \\ &= \frac{1}{2}\Big(\sin(2x) - \frac{1}{3}\sin^3(2x)\Big) + C\end{aligned}$$

Putting all the pieces together, we have

$$\begin{aligned}\int \cos^4 x \sin^2 x\ dx &= \int \frac{1}{8}\big(1+\cos(2x)-\cos^2(2x)-\cos^3(2x)\big)\ dx\\ &= \frac{1}{8}\Big[x+\frac{1}{2}\sin(2x)-\frac{1}{2}\Big(x+\frac{1}{4}\sin(4x)\Big)-\frac{1}{2}\Big(\sin(2x)-\frac{1}{3}\sin^3(2x)\Big)\Big]+C\\ &= \frac{1}{8}\Big[\frac{1}{2}x-\frac{1}{8}\sin(4x)+\frac{1}{6}\sin^3(2x)\Big]+C\end{aligned}$$

The process above was a bit long and tedious, but being able to work a problem such as this from start to finish is important.

Integrals of the form $\int \sin(mx)\sin(nx)\ dx$, $\int \cos(mx)\cos(nx)\ dx$, and $\int \sin(mx)\cos(nx)\ dx$.

Functions that contain products of sines and cosines of differing periods are important in many applications including the analysis of sound waves. Integrals of the form

$$\int \sin(mx)\sin(nx)\ dx,\quad \int \cos(mx)\cos(nx)\ dx \quad \text{and} \quad \int \sin(mx)\cos(nx)\ dx \tag{6.3.8}$$

are best approached by first applying the Product to Sum Formulas found in the back cover of this text, namely

$$\begin{aligned}\sin(mx)\sin(nx) &= \frac{1}{2}\Big[\cos\big((m-n)x\big)-\cos\big((m+n)x\big)\Big]\\ \cos(mx)\cos(nx) &= \frac{1}{2}\Big[\cos\big((m-n)x\big)+\cos\big((m+n)x\big)\Big]\\ \sin(mx)\cos(nx) &= \frac{1}{2}\Big[\sin\big((m-n)x\big)+\sin\big((m+n)x\big)\Big]\end{aligned}$$

Example 6.3.4: Integrating products of $\sin(mx)$ and $\cos(nx)$

Evaluate $\int \sin(5x)\cos(2x)\ dx$.

Solution

The application of the formula and subsequent integration are straightforward:

$$\begin{aligned}\int \sin(5x)\cos(2x)\ dx &= \int \frac{1}{2}\Big[\sin(3x)+\sin(7x)\Big]\ dx\\ &= -\frac{1}{6}\cos(3x)-\frac{1}{14}\cos(7x)+C\end{aligned}$$

Integrals of the form $\int \tan^m x \sec^n x\ dx$.

When evaluating integrals of the form $\int \sin^m x \cos^n x\ dx$, the Pythagorean Theorem allowed us to convert even powers of sine into even powers of cosine, and vise--versa. If, for instance, the power of sine was odd, we pulled out one $\sin x$ and converted the remaining even power of $\sin x$ into a function using powers of $\cos x$, leading to an easy substitution.

The same basic strategy applies to integrals of the form $\int \tan^m x \sec^n x\ dx$, albeit a bit more nuanced. The following three facts will prove useful:

1. $\frac{d}{dx}(\tan x) = \sec^2 x$,
2. $\frac{d}{dx}(\sec x) = \sec x \tan x$, and
3. $1+\tan^2 x = \sec^2 x$ (the Pythagorean Theorem).

If the integrand can be manipulated to separate a $\sec^2 x$ term with the remaining secant power even, or if a $\sec x \tan x$ term can be separated with the remaining $\tan x$ power even, the Pythagorean Theorem can be employed, leading to a simple substitution. This strategy is outlined in the following Key Idea.

Key Idea 12: Integrals Involving Powers of Tangent and Secant

Consider $\int \tan^m x \sec^n x\ dx$, where m, n are nonnegative integers.

1. If n is even, then $n = 2k$ for some integer k. Rewrite $\sec^n x$ as

$$\sec^n x = \sec^{2k} x = \sec^{2k-2} x \sec^2 x = (1+\tan^2 x)^{k-1}\sec^2 x. \tag{6.3.9}$$

Then

$$\begin{aligned}\int \tan^m x \sec^n x \; dx &= \int \tan^m x(1+\tan^2 x)^{k-1} \sec^2 x \; dx \\ &= \int u^m(1+u^2)^{k-1} \; du,\end{aligned} \tag{6.3.10}$$

where $u = \tan x$ and $du = \sec^2 x \; dx$.

2. If m is odd, then $m = 2k+1$ for some integer k. Rewrite $\tan^m x \sec^n x$ as

$$\begin{aligned}\tan^m x \sec^n x = \tan^{2k+1} x \sec^n x &= \tan^{2k} x \sec^{n-1} x \sec x \tan x \\ &= (\sec^2 x - 1)^k \sec^{n-1} x \sec x \tan x.\end{aligned} \tag{6.3.11}$$

Then

$$\begin{aligned}\int \tan^m x \sec^n x \; dx &= \int (\sec^2 x - 1)^k \sec^{n-1} x \sec x \tan x \; dx \\ &= \int (u^2-1)^k u^{n-1} \; du,\end{aligned} \tag{6.3.12}$$

where $u = \sec x$ and $du = \sec x \tan x \; dx$.

3. If n is odd and m is even, then $m = 2k$ for some integer k. Convert $\tan^m x$ to $(\sec^2 x - 1)^k$. Expand the new integrand and use Integration By Parts, with $dv = \sec^2 x \; dx$.
4. If m is even and $n = 0$, rewrite $\tan^m x$ as

$$\tan^m x = \tan^{m-2} x \tan^2 x = \tan^{m-2} x(\sec^2 x - 1) = \tan^{m-2} \sec^2 x - \tan^{m-2} x. \tag{6.3.13}$$

So

$$\int \tan^m x \; dx = \underbrace{\int \tan^{m-2} \sec^2 x \; dx}_{\text{apply rule \#1}} \quad - \underbrace{\int \tan^{m-2} x \; dx}_{\text{apply rule \#4 again}}. \tag{6.3.14}$$

The techniques described in items 1 and 2 of Key Idea 12 are relatively straightforward, but the techniques in items 3 and 4 can be rather tedious. A few examples will help with these methods.

Example 6.3.5: Integrating powers of tangent and secant

Evaluate $\int \tan^2 x \sec^6 x \; dx$.

Solution

Since the power of secant is even, we use rule #1 from Key Idea 12 and pull out a $\sec^2 x$ in the integrand. We convert the remaining powers of secant into powers of tangent.

$$\begin{aligned}\int \tan^2 x \sec^6 x \; dx &= \int \tan^2 x \sec^4 x \sec^2 x \; dx \\ &= \int \tan^2 x \left(1+\tan^2 x\right)^2 \sec^2 x \; dx\end{aligned}$$

Now substitute, with $u = \tan x$, with $du = \sec^2 x \; dx$.

$$= \int u^2 \left(1+u^2\right)^2 \; du$$

We leave the integration and subsequent substitution to the reader. The final answer is

$$= \frac{1}{3}\tan^3 x + \frac{2}{5}\tan^5 x + \frac{1}{7}\tan^7 x + C.$$

Example 6.3.6: Integrating powers of tangent and secant

Evaluate $\int \sec^3 x\ dx$.

Solution

We apply rule #3 from Key Idea 12 as the power of secant is odd and the power of tangent is even (0 is an even number). We use Integration by Parts; the rule suggests letting $dv = \sec^2 x\ dx$, meaning that $u = \sec x$.

$$\begin{array}{ll} u = \sec x & v = ? \\ du = ? & dv = \sec^2 x\ dx \end{array}$$

***Figure** 6.3.2**:** Setting up Integration by Parts.*

Employing Integration by Parts, we have

$$\begin{aligned} \int \sec^3 x\ dx &= \int \underbrace{\sec x}_{u} \cdot \underbrace{\sec^2 x\ dx}_{dv} \\ &= \sec x \tan x - \int \sec x \tan^2 x\ dx. \end{aligned}$$

This new integral also requires applying rule \#3 of Key Idea:

$$\begin{aligned} &= \sec x \tan x - \int \sec x \left(\sec^2 x - 1\right)\ dx \\ &= \sec x \tan x - \int \sec^3 x\ dx + \int \sec x\ dx \\ &= \sec x \tan x - \int \sec^3 x\ dx + \ln|\sec x + \tan x| \end{aligned}$$

In previous applications of Integration by Parts, we have seen where the original integral has reappeared in our work. We resolve this by adding $\int \sec^3 x\ dx$ to both sides, giving:

$$\begin{aligned} 2\int \sec^3 x\ dx &= \sec x \tan x + \ln|\sec x + \tan x| \\ \int \sec^3 x\ dx &= \frac{1}{2}\left(\sec x \tan x + \ln|\sec x + \tan x|\right) + C \end{aligned}$$

We give one more example.

Example 6.3.7: Integrating powers of tangent and secant

Evaluate $\int \tan^6 x\ dx$.

Solution

We employ rule #4 of Key Idea 12.

$$\begin{aligned} \int \tan^6 x\ dx &= \int \tan^4 x \tan^2 x\ dx \\ &= \int \tan^4 x\left(\sec^2 x - 1\right)\ dx \\ &= \int \tan^4 x \sec^2 x\ dx - \int \tan^4 x\ dx \end{aligned}$$

Integrate the first integral with substitution, $u = \tan x$; integrate the second by employing rule #4 again.

$$\begin{aligned} &= \frac{1}{5}\tan^5 x - \int \tan^2 x \tan^2 x\ dx \\ &= \frac{1}{5}\tan^5 x - \int \tan^2 x\left(\sec^2 x - 1\right)\ dx \\ &= \frac{1}{5}\tan^5 x - \int \tan^2 x \sec^2 x\ dx + \int \tan^2 x\ dx \end{aligned}$$

> Again, use substitution for the first integral and rule \#4 for the second.
>
> $$\begin{aligned} &= \frac{1}{5}\tan^5 x - \frac{1}{3}\tan^3 x + \int \left(\sec^2 x - 1\right)\,dx \\ &= \frac{1}{5}\tan^5 x - \frac{1}{3}\tan^3 x + \tan x - x + C. \end{aligned}$$

These latter examples were admittedly long, with repeated applications of the same rule. Try to not be overwhelmed by the length of the problem, but rather admire how robust this solution method is. A trigonometric function of a high power can be systematically reduced to trigonometric functions of lower powers until all antiderivatives can be computed.

The next section introduces an integration technique known as Trigonometric Substitution, a clever combination of Substitution and the Pythagorean Theorem.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/
- Integrated by Justin Marshall.

This page titled 6.3: Trigonometric Integrals is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

6.4: Trigonometric Substitution

In Section 5.2 we defined the definite integral as the signed area under the curve." In that section we had not yet learned the Fundamental Theorem of Calculus, so we evaluated special definite integrals which described nice, geometric shapes. For instance, we were able to evaluate

$$\int_{-3}^{3}\sqrt{9-x^2}\,dx=\frac{9\pi}{2} \tag{6.4.1}$$

as we recognized that $f(x)=\sqrt{9-x^2}$ described the upper half of a circle with radius 3.

We have since learned a number of integration techniques, including Substitution and Integration by Parts, yet we are still unable to evaluate the above integral without resorting to a geometric interpretation. This section introduces Trigonometric Substitution, a method of integration that fills this gap in our integration skill. This technique works on the same principle as Substitution as found in Section 6.1, though it can feel "backward." In Section 6.1, we set $u=f(x)$, for some function f, and replaced $f(x)$ with u. In this section, we will set $x=f(\theta)$, where f is a trigonometric function, then replace x with $f(\theta)$.

We start by demonstrating this method in evaluating the integral in $6.4.1$. After the example, we will generalize the method and give more examples.

Example $6.4.1$: Using Trigonometric Substitution

Evaluate $\int_{-3}^{3}\sqrt{9-x^2}\,dx$.

Solution

We begin by noting that $9\sin^2\theta+9\cos^2\theta=9$, and hence $9\cos^2\theta=9-9\sin^2\theta$. If we let $x=3\sin\theta$, then $9-x^2=9-9\sin^2\theta=9\cos^2\theta$.

Setting $x=3\sin\theta$ gives $dx=3\cos\theta\,d\theta$. We are almost ready to substitute. We also wish to change our bounds of integration. The bound $x=-3$ corresponds to $\theta=-\pi/2$ (for when $\theta=-\pi/2$, $x=3\sin\theta=-3$). Likewise, the bound of $x=3$ is replaced by the bound $\theta=\pi/2$. Thus

$$\begin{aligned}\int_{-3}^{3}\sqrt{9-x^2}\,dx &= \int_{-\pi/2}^{\pi/2}\sqrt{9-9\sin^2\theta}(3\cos\theta)\,d\theta &(6.4.2)\\ &= \int_{-\pi/2}^{\pi/2}3\sqrt{9\cos^2\theta}\cos\theta\,d\theta &(6.4.3)\\ &= \int_{-\pi/2}^{\pi/2}3|3\cos\theta|\cos\theta\,d\theta. &(6.4.4)\end{aligned}$$

On $[-\pi/2,\pi/2]$, $\cos\theta$ is always positive, so we can drop the absolute value bars, then employ a power--reducing formula:

$$\begin{aligned}&= \int_{-\pi/2}^{\pi/2}9\cos^2\theta\,d\theta &(6.4.5)\\ &= \int_{-\pi/2}^{\pi/2}\frac{9}{2}\big(1+\cos(2\theta)\big)\,d\theta &(6.4.6)\\ &= \frac{9}{2}\left(\theta+\frac{1}{2}\sin(2\theta)\right)\Big|_{-\pi/2}^{\pi/2}=\frac{9}{2}\pi. &(6.4.7)\end{aligned}$$

This matches our answer from before.

We now describe in detail Trigonometric Substitution. This method excels when dealing with integrands that contain $\sqrt{a^2-x^2}$, $\sqrt{x^2-a^2}$ and $\sqrt{x^2+a^2}$. The following Key Idea 13 outlines the procedure for each case, followed by more examples. Each right triangle acts as a reference to help us understand the relationships between x and θ.

Key Idea 13: Trigonometric Substitution

a. For integrands containing $\sqrt{a^2-x^2}$:
   Let $x=a\sin\theta$, $dx=a\cos\theta\,d\theta$
   Thus $\theta=\sin^{-1}(x/a)$, for $-\pi/2\le\theta\le\pi/2$.

On this interval, $\cos\theta \geq 0$, so
$\sqrt{a^2-x^2} = a\cos\theta$.

b. For integrands containing $\sqrt{x^2+a^2}$:
Let $x = a\tan\theta$, $dx = a\sec^2\theta\ d\theta$
Thus $\theta = \tan^{-1}(x/a)$, for $-\pi/2 < \theta < \pi/2$.
On this interval, $\sec\theta > 0$, so
$\sqrt{x^2+a^2} = a\sec\theta$.

c. For integrands containing $\sqrt{x^2-a^2}$:
Let $x = a\sec\theta$, $dx = a\sec\theta\tan\theta\ d\theta$
Thus $\theta = \sec^{-1}(x/a)$. If $x/a \geq 1$, then $0 \leq \theta < \pi/2$; if $x/a \leq -1$, then $\pi/2 < \theta \leq \pi$.
We restrict our work to where $x \geq a$, so $x/a \geq 1$, and $0 \leq \theta < \pi/2$.
On this interval, $\tan\theta \geq 0$, so
$\sqrt{x^2-a^2} = a\tan\theta$.

Example 6.4.2: Using Trigonometric Substitution

Evaluate $\int \frac{1}{\sqrt{5+x^2}}\ dx$.

Solution

Using Key Idea 13(b), we recognize $a = \sqrt{5}$ and set $x = \sqrt{5}\tan\theta$. This makes $dx = \sqrt{5}\sec^2\theta\ d\theta$. We will use the fact that $\sqrt{5+x^2} = \sqrt{5+5\tan^2\theta} = \sqrt{5\sec^2\theta} = \sqrt{5}\sec\theta$. Substituting, we have:

$$\begin{aligned}\int \frac{1}{\sqrt{5+x^2}}\ dx &= \int \frac{1}{\sqrt{5+5\tan^2\theta}}\sqrt{5}\sec^2\theta\ d\theta \qquad (6.4.8)\\ &= \int \frac{\sqrt{5}\sec^2\theta}{\sqrt{5}\sec\theta}\ d\theta \qquad (6.4.9)\\ &= \int \sec\theta\ d\theta \qquad (6.4.10)\\ &= \ln\left|\sec\theta+\tan\theta\right| + C. \qquad (6.4.11)\end{aligned}$$

While the integration steps are over, we are not yet done. The original problem was stated in terms of x, whereas our answer is given in terms of θ. We must convert back to x.

The reference triangle given in Key Idea 13(b) helps. With $x = \sqrt{5}\tan\theta$, we have

$$\tan\theta = \frac{x}{\sqrt{5}} \quad \text{and} \quad \sec\theta = \frac{\sqrt{x^2+5}}{\sqrt{5}}. \qquad (6.4.12)$$

This gives

$$\begin{aligned}\int \frac{1}{\sqrt{5+x^2}}\ dx &= \ln\left|\sec\theta+\tan\theta\right| + C \qquad (6.4.13)\\ &= \ln\left|\frac{\sqrt{x^2+5}}{\sqrt{5}} + \frac{x}{\sqrt{5}}\right| + C. \qquad (6.4.14)\end{aligned}$$

We can leave this answer as is, or we can use a logarithmic identity to simplify it. Note:

$$\begin{aligned}\ln\left|\frac{\sqrt{x^2+5}}{\sqrt{5}} + \frac{x}{\sqrt{5}}\right| + C &= \ln\left|\frac{1}{\sqrt{5}}\left(\sqrt{x^2+5}+x\right)\right| + C \qquad (6.4.15)\\ &= \ln\left|\frac{1}{\sqrt{5}}\right| + \ln\left|\sqrt{x^2+5}+x\right| + C \qquad (6.4.16)\\ &= \ln\left|\sqrt{x^2+5}+x\right| + C, \qquad (6.4.17)\end{aligned}$$

where the $\ln\left(1/\sqrt{5}\right)$ term is absorbed into the constant C. (In Section 6.6 we will learn another way of approaching this problem.)

Example $6.4.3$: Using Trigonometric Substitution

Evaluate $\int \sqrt{4x^2-1}\ dx$.

Solution

We start by rewriting the integrand so that it looks like $\sqrt{x^2-a^2}$ for some value of a:

$$\sqrt{4x^2-1} = \sqrt{4\left(x^2-\frac{1}{4}\right)} \tag{6.4.18}$$

$$= 2\sqrt{x^2-\left(\frac{1}{2}\right)^2} \tag{6.4.19}$$

So we have $a = 1/2$, and following Key Idea 13(c), we set $x = \frac{1}{2}\sec\theta$, and hence $dx = \frac{1}{2}\sec\theta\tan\theta\ d\theta$.

We now rewrite the integral with these substitutions:

$$\int \sqrt{4x^2-1}\ dx = \int 2\sqrt{x^2-\left(\frac{1}{2}\right)^2}\ dx \tag{6.4.20}$$

$$= \int 2\sqrt{\frac{1}{4}\sec^2\theta-\frac{1}{4}}\left(\frac{1}{2}\sec\theta\tan\theta\right)\ d\theta \tag{6.4.21}$$

$$= \int \sqrt{\frac{1}{4}(\sec^2\theta-1)}\Big(\sec\theta\tan\theta\Big)\ d\theta \tag{6.4.22}$$

$$= \int \sqrt{\frac{1}{4}\tan^2\theta}\Big(\sec\theta\tan\theta\Big)\ d\theta \tag{6.4.23}$$

$$= \int \frac{1}{2}\tan^2\theta\sec\theta\ d\theta \tag{6.4.24}$$

$$= \frac{1}{2}\int\Big(\sec^2\theta-1\Big)\sec\theta\ d\theta \tag{6.4.25}$$

$$= \frac{1}{2}\int\Big(\sec^3\theta-\sec\theta\Big)\ d\theta. \tag{6.4.26}$$

We integrated $\sec^3\theta$ in Example 6.3.6, finding its antiderivatives to be

$$\int \sec^3\theta\ d\theta = \frac{1}{2}\Big(\sec\theta\tan\theta+\ln|\sec\theta+\tan\theta|\Big)+C. \tag{6.4.27}$$

Thus

$$\int \sqrt{4x^2-1}\ dx = \frac{1}{2}\int\Big(\sec^3\theta-\sec\theta\Big)\ d\theta \tag{6.4.28}$$

$$= \frac{1}{2}\left(\frac{1}{2}\Big(\sec\theta\tan\theta+\ln|\sec\theta+\tan\theta|\Big) - \ln|\sec\theta+\tan\theta|\right)+C \tag{6.4.29}$$

$$= \frac{1}{4}\left(\sec\theta\tan\theta-\ln|\sec\theta+\tan\theta|\right)+C. \tag{6.4.30}$$

We are not yet done. Our original integral is given in terms of x, whereas our final answer, as given, is in terms of θ. We need to rewrite our answer in terms of x. With $a = 1/2$, and $x = \frac{1}{2}\sec\theta$, the reference triangle in Key Idea 13(c) shows that

$$\tan\theta = \sqrt{x^2-1/4}\Big/(1/2) = 2\sqrt{x^2-1/4} \quad \text{and} \quad \sec\theta = 2x. \tag{6.4.31}$$

Thus

$$\frac{1}{4}\Big(\sec\theta\tan\theta-\ln\big|\sec\theta+\tan\theta\big|\Big)+C = \frac{1}{4}\Big(2x\cdot 2\sqrt{x^2-1/4} - \ln\big|2x+2\sqrt{x^2-1/4}\big|\Big)+C \tag{6.4.32}$$

$$= \frac{1}{4}\Big(4x\sqrt{x^2-1/4} - \ln\big|2x+2\sqrt{x^2-1/4}\big|\Big)+C. \tag{6.4.33}$$

The final answer is given in the last line above, repeated here:

$$\int \sqrt{4x^2-1}\;dx = \frac{1}{4}\Big(4x\sqrt{x^2-1/4} - \ln\left|2x+2\sqrt{x^2-1/4}\right|\Big)+C. \tag{6.4.34}$$

Example 6.4.4: Using Trigonometric Substitution

Evaluate $\int \frac{\sqrt{4-x^2}}{x^2}\;dx$.

Solution

We use Key Idea 13(a) with $a=2$, $x=2\sin\theta$, $dx=2\cos\theta$ and hence $\sqrt{4-x^2}=2\cos\theta$. This gives

$$\begin{aligned}\int\frac{\sqrt{4-x^2}}{x^2}\;dx &= \int\frac{2\cos\theta}{4\sin^2\theta}(2\cos\theta)\;d\theta &(6.4.35)\\ &=\int\cot^2\theta\;d\theta &(6.4.36)\\ &=\int(\csc^2\theta-1)\;d\theta &(6.4.37)\\ &=-\cot\theta-\theta+C. &(6.4.38)\end{aligned}$$

We need to rewrite our answer in terms of x. Using the reference triangle found in Key Idea 13(a), we have $\cot\theta=\sqrt{4-x^2}/x$ and $\theta=\sin^{-1}(x/2)$. Thus

$$\int\frac{\sqrt{4-x^2}}{x^2}\;dx = -\frac{\sqrt{4-x^2}}{x}-\sin^{-1}\left(\frac{x}{2}\right)+C. \tag{6.4.39}$$

Trigonometric Substitution can be applied in many situations, even those not of the form $\sqrt{a^2-x^2}$, $\sqrt{x^2-a^2}$ or $\sqrt{x^2+a^2}$. In the following example, we apply it to an integral we already know how to handle.

Example 6.4.5: Using Trigonometric Substitution

Evaluate $\int\frac{1}{x^2+1}\;dx$.

Solution

We know the answer already as $\tan^{-1}x+C$. We apply Trigonometric Substitution here to show that we get the same answer without inherently relying on knowledge of the derivative of the arctangent function.

Using Key Idea 13(b), let $x=\tan\theta$, $dx=\sec^2\theta\;d\theta$ and note that $x^2+1=\tan^2\theta+1=\sec^2\theta$. Thus

$$\begin{aligned}\int\frac{1}{x^2+1}\;dx &= \int\frac{1}{\sec^2\theta}\sec^2\theta\;d\theta &(6.4.40)\\ &=\int 1\;d\theta &(6.4.41)\\ &=\theta+C. &(6.4.42)\end{aligned}$$

Since $x=\tan\theta$, $\theta=\tan^{-1}x$, and we conclude that $\int\frac{1}{x^2+1}\;dx=\tan^{-1}x+C$.

The next example is similar to the previous one in that it does not involve a square--root. It shows how several techniques and identities can be combined to obtain a solution.

Example 6.4.6: Using Trigonometric Substitution

Evaluate $\int\frac{1}{(x^2+6x+10)^2}\;dx$.

Solution

We start by completing the square, then make the substitution $u=x+3$, followed by the trigonometric substitution of $u=\tan\theta$:

$$\int\frac{1}{(x^2+6x+10)^2}\;dx=\int\frac{1}{\big((x+3)^2+1\big)^2}\;dx=\int\frac{1}{(u^2+1)^2}\;du. \tag{6.4.43}$$

Now make the substitution $u=\tan\theta$, $du=\sec^2\theta\;d\theta$:

$$= \int \frac{}{(\tan^2\theta+1)^2}\sec^2\theta\ d\theta \tag{6.4.44}$$

$$= \int \frac{1}{(\sec^2\theta)^2}\sec^2\theta\ d\theta \tag{6.4.45}$$

$$= \int \cos^2\theta\ d\theta. \tag{6.4.46}$$

Applying a power reducing formula, we have

$$= \int \left(\frac{1}{2}+\frac{1}{2}\cos(2\theta)\right)\ d\theta \tag{6.4.47}$$

$$= \frac{1}{2}\theta+\frac{1}{4}\sin(2\theta)+C. \tag{6.4.48}$$

We need to return to the variable x. As $u=\tan\theta$, $\theta=\tan^{-1}u$. Using the identity $\sin(2\theta)=2\sin\theta\cos\theta$ and using the reference triangle found in Key Idea 13(b), we have

$$\frac{1}{4}\sin(2\theta)=\frac{1}{2}\frac{u}{\sqrt{u^2+1}}\cdot\frac{1}{\sqrt{u^2+1}}=\frac{1}{2}\frac{u}{u^2+1}. \tag{6.4.49}$$

Finally, we return to x with the substitution $u=x+3$. We start with the expression in Equation (???):

$$\frac{1}{2}\theta+\frac{1}{4}\sin(2\theta)+C=\frac{1}{2}\tan^{-1}u+\frac{1}{2}\frac{u}{u^2+1}+C \tag{6.4.50}$$

$$=\frac{1}{2}\tan^{-1}(x+3)+\frac{x+3}{2(x^2+6x+10)}+C. \tag{6.4.51}$$

Stating our final result in one line,

$$\int \frac{1}{(x^2+6x+10)^2}\ dx=\frac{1}{2}\tan^{-1}(x+3)+\frac{x+3}{2(x^2+6x+10)}+C. \tag{6.4.52}$$

Our last example returns us to definite integrals, as seen in our first example. Given a definite integral that can be evaluated using Trigonometric Substitution, we could first evaluate the corresponding indefinite integral (by changing from an integral in terms of x to one in terms of θ, then converting back to x) and then evaluate using the original bounds. It is much more straightforward, though, to change the bounds as we substitute.

Example $6.4.7$: Definite integration and Trigonometric Substitution

Evaluate $\int_0^5 \frac{x^2}{\sqrt{x^2+25}}\ dx$.

Solution

Using Key Idea 13(b), we set $x=5\tan\theta$, $dx=5\sec^2\theta\ d\theta$, and note that $\sqrt{x^2+25}=5\sec\theta$. As we substitute, we can also change the bounds of integration.

The lower bound of the original integral is $x=0$. As $x=5\tan\theta$, we solve for θ and find $\theta=\tan^{-1}(x/5)$. Thus the new lower bound is $\theta=\tan^{-1}(0)=0$. The original upper bound is $x=5$, thus the new upper bound is $\theta=\tan^{-1}(5/5)=\pi/4$.

Thus we have

$$\int_0^5 \frac{x^2}{\sqrt{x^2+25}}\ dx=\int_0^{\pi/4}\frac{25\tan^2\theta}{5\sec\theta}5\sec^2\theta\ d\theta \tag{6.4.53}$$

$$=25\int_0^{\pi/4}\tan^2\theta\sec\theta\ d\theta. \tag{6.4.54}$$

We encountered this indefinite integral in Example $6.4.3$ where we found

$$\int \tan^2\theta\sec\theta\ d\theta=\frac{1}{2}\Big(\sec\theta\tan\theta-\ln|\sec\theta+\tan\theta|\Big). \tag{6.4.55}$$

So

$$25\int_0^{\pi/4}\tan^2\theta\sec\theta\ d\theta=\frac{25}{2}\Big(\sec\theta\tan\theta-\ln|\sec\theta+\tan\theta|\Big)\Big|_0^{\pi/4} \tag{6.4.56}$$

$$= \frac{\quad}{2}\left(\sqrt{2}-\ln(\sqrt{2}+1)\right) \tag{6.4.57}$$

$$\approx 6.661. \tag{6.4.58}$$

The following equalities are very useful when evaluating integrals using Trigonometric Substitution.

Ket Idea 14: Useful Equalities with Trigonometric Substitution

1. $\sin(2\theta) = 2\sin\theta\cos\theta$
2. $\cos(2\theta) = \cos^2\theta - \sin^2\theta = 2\cos^2\theta - 1 = 1 - 2\sin^2\theta$
3. $\int \sec^3\theta \; d\theta = \frac{1}{2}\Big(\sec\theta\tan\theta + \ln\left|\sec\theta + \tan\theta\right|\Big) + C$
4. $\int \cos^2\theta \; d\theta = \int \frac{1}{2}\big(1 + \cos(2\theta)\big) \; d\theta = \frac{1}{2}\big(\theta + \sin\theta\cos\theta\big) + C.$

The next section introduces Partial Fraction Decomposition, which is an algebraic technique that turns "complicated" fractions into sums of "simpler" fractions, making integration easier.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/
- Integrated by Justin Marshall.

This page titled 6.4: Trigonometric Substitution is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

6.5: Partial Fraction Decomposition

In this section we investigate the antiderivatives of rational functions. Recall that rational functions are functions of the form $f(x) = \frac{p(x)}{q(x)}$, where $p(x)$ and $q(x)$ are polynomials and $q(x) \neq 0$. Such functions arise in many contexts, one of which is the solving of certain fundamental differential equations.

We begin with an example that demonstrates the motivation behind this section. Consider the integral $\int \frac{1}{x^2-1}\, dx$. We do not have a simple formula for this (if the denominator were x^2+1, we would recognize the antiderivative as being the arctangent function). It can be solved using Trigonometric Substitution, but note how the integral is easy to evaluate once we realize:

$\frac{1}{x^2-1} = \frac{1/2}{x-1} - \frac{1/2}{x+1}$

Thus

$$\int \frac{1}{x^2-1}\, dx = \int \frac{1/2}{x-1}\, dx - \int \frac{1/2}{x+1}\, dx \tag{6.5.1}$$

$$= \frac{1}{2}\ln|x-1| - \frac{1}{2}\ln|x+1| + C. \tag{6.5.2}$$

This section teaches how to *decompose*

$\frac{1}{x^2-1}$ into $\frac{1/2}{x-1} - \frac{1/2}{x+1}$

We start with a rational function $f(x) = \frac{p(x)}{q(x)}$, where p and q do not have any common factors and the degree of p is less than the degree of q. It can be shown that any polynomial, and hence q, can be factored into a product of linear and irreducible quadratic terms. The following Key Idea states how to decompose a rational function into a sum of rational functions whose denominators are all of lower degree than q.

Key Idea 15: Partial Fraction Decomposition

Let $\frac{p(x)}{q(x)}$ be a rational function, where the degree of p is less than the degree of q.

1. **Linear Terms:** Let $(x-a)$ divide $q(x)$, where $(x-a)^n$ is the highest power of $(x-a)$ that divides $q(x)$. Then the decomposition of $\frac{p(x)}{q(x)}$ will contain the sum

$$\frac{A_1}{(x-a)} + \frac{A_2}{(x-a)^2} + \cdots + \frac{A_n}{(x-a)^n}. \tag{6.5.3}$$

2. **Quadratic Terms:** Let x^2+bx+c divide $q(x)$, where $(x^2+bx+c)^n$ is the highest power of x^2+bx+c that divides $q(x)$. Then the decomposition of $\frac{p(x)}{q(x)}$ will contain the sum

$$\frac{B_1x+C_1}{x^2+bx+c} + \frac{B_2x+C_2}{(x^2+bx+c)^2} + \cdots + \frac{B_nx+C_n}{(x^2+bx+c)^n}. \tag{6.5.4}$$

To find the coefficients A_i, B_i and C_i:

1. Multiply all fractions by $q(x)$, clearing the denominators. Collect like terms.
2. Equate the resulting coefficients of the powers of x and solve the resulting system of linear equations.

The following examples will demonstrate how to put this Key Idea into practice. Example $6.5.1$ stresses the decomposition aspect of the Key Idea.

Example $6.5.1$: Decomposing into partial fractions

Decompose $f(x) = \frac{1}{(x+5)(x-2)^3(x^2+x+2)(x^2+x+7)^2}$ without solving for the resulting coefficients.

Solution

The denominator is already factored, as both x^2+x+2 and x^2+x+7 cannot be factored further. We need to decompose $f(x)$ properly. Since $(x+5)$ is a linear term that divides the denominator, there will be a

$\frac{A}{x+5}$

term in the decomposition.

As $(x-2)^3$ divides the denominator, we will have the following terms in the decomposition:

$$\frac{B}{x-2},\quad \frac{C}{(x-2)^2}\quad \text{and}\quad \frac{D}{(x-2)^3}.\]

The x^2+x+2 term in the denominator results in a $\frac{Ex+F}{x^2+x+2}$ term.

Finally, the $(x^2+x+7)^2$ term results in the terms

$$\frac{Gx+H}{x^2+x+7}\quad \text{and}\quad \frac{Ix+J}{(x^2+x+7)^2}.\]

All together, we have

$$\begin{aligned} \frac{1}{(x+5)(x-2)^3(x^2+x+2)(x^2+x+7)^2} &= \frac{A}{x+5}+\frac{B}{x-2}+\frac{C}{(x-2)^2} \\ &\quad +\frac{D}{(x-2)^3}+ \qquad (6.5.5) \\ &\frac{Ex+F}{x^2+x+2}+\frac{Gx+H}{x^2+x+7} \\ &\quad +\frac{Ix+J}{(x^2+x+7)^2} \qquad (6.5.6) \end{aligned}$$

Solving for the coefficients A, $B \ldots J$ would be a bit tedious but not "hard."

Example 6.5.2: Decomposing into partial fractions

Perform the partial fraction decomposition of $\frac{1}{x^2-1}$.

Solution

The denominator factors into two linear terms: $x^2-1=(x-1)(x+1)$. Thus

$$\frac{1}{x^2-1} = \frac{A}{x-1} + \frac{B}{x+1}.\]

To solve for A and B, first multiply through by $x^2-1=(x-1)(x+1)$:

$$\begin{aligned} 1 &= \frac{A(x-1)(x+1)}{x-1}+\frac{B(x-1)(x+1)}{x+1} \qquad (6.5.7) \\ &= A(x+1)+B(x-1) \qquad (6.5.8) \\ &= Ax+A+Bx-B \qquad (6.5.9) \end{aligned}$$

Now collect like terms.

$$=(A+B)x+(A-B). \qquad (6.5.10)$$

The next step is key. Note the equality we have:

$$1=(A+B)x+(A-B). \qquad (6.5.11)$$

For clarity's sake, rewrite the left hand side as

$$0x+1=(A+B)x+(A-B). \qquad (6.5.12)$$

On the left, the coefficient of the x term is 0; on the right, it is $(A+B)$. Since both sides are equal, we must have that $0=A+B$.

Likewise, on the left, we have a constant term of 1; on the right, the constant term is $(A-B)$. Therefore we have $1=A-B$.

We have two linear equations with two unknowns. This one is easy to solve by hand, leading to

$$\begin{array}{r} A+B=0 \\ A-B=1 \end{array} \quad \Rightarrow \quad \begin{array}{r} A=1/2 \\ B=-1/2 \end{array}. \qquad (6.5.13)$$

Thus $$\frac{1}{x^2-1} = \frac{1/2}{x-1}-\frac{1/2}{x+1}.\]

Example $6.5.3$: Integrating using partial fractions

Use partial fraction decomposition to integrate $\int \frac{1}{(x-1)(x+2)^2}\,dx$.

Solution

We decompose the integrand as follows, as described by Key Idea 15:

$$\frac{1}{(x-1)(x+2)^2} = \frac{A}{x-1} + \frac{B}{x+2} + \frac{C}{(x+2)^2}.\]

To solve for A, B and C, we multiply both sides by $(x-1)(x+2)^2$ and collect like terms:

$$\begin{aligned} 1 &= A(x+2)^2 + B(x-1)(x+2) + C(x-1) && (6.5.14) \\ &= Ax^2 + 4Ax + 4A + Bx^2 + Bx - 2B + Cx - C && (6.5.15) \\ &= (A+B)x^2 + (4A+B+C)x + (4A-2B-C) && (6.5.16) \end{aligned}$$

Note: Equation $6.5.22$ offers a direct route to finding the values of A, B and C. Since the equation holds for all values of x, it holds in particular when $x=1$. However, when $x=1$, the right hand side simplifies to $A(1+2)^2 = 9A$. Since the left hand side is still 1, we have $1 = 9A$. Hence $A = 1/9$.

Likewise, the equality holds when $x=-2$; this leads to the equation $1 = -3C$. Thus $C = -1/3$.

Knowing A and C, we can find the value of B by choosing yet another value of x, such as $x=0$, and solving for B.

We have

$$0x^2+0x+ 1 = (A+B)x^2 + (4A+B+C)x + (4A-2B-C)\]

leading to the equations

$$A+B = 0, \quad 4A+B+C = 0 \quad \text{and} \quad 4A-2B-C = 1.\]

These three equations of three unknowns lead to a unique solution:

$$A = 1/9,\quad B = -1/9 \quad \text{and} \quad C = -1/3.\]

Thus

$$\int\frac{1}{(x-1)(x+2)^2}\ dx = \int \frac{1/9}{x-1}\ dx + \int \frac{-1/9}{x+2}\ dx + \int \frac{-1/3}{(x+2)^2}\ dx.\]

Each can be integrated with a simple substitution with $u = x-1$ or $u = x+2$ (or by directly applying Key Idea 10 as the denominators are linear functions). The end result is

$$\int\frac{1}{(x-1)(x+2)^2}\ dx = \frac19\ln|x-1| -\frac19\ln|x+2| +\frac1{3(x+2)}+C.\]

Example $6.5.4$: Integrating using partial fractions

Use partial fraction decomposition to integrate $\int \frac{x^3}{(x-5)(x+3)}\,dx$.

Solution

Key Idea 15 presumes that the degree of the numerator is less than the degree of the denominator. Since this is not the case here, we begin by using polynomial division to reduce the degree of the numerator. We omit the steps, but encourage the reader to verify that

$$\frac{x^3}{(x-5)(x+3)} = x+2+\frac{19x+30}{(x-5)(x+3)}.\]

Using Key Idea 15, we can rewrite the new rational function as:

$$\frac{19x+30}{(x-5)(x+3)} = \frac{A}{x-5} + \frac{B}{x+3}\]

for appropriate values of A and B. Clearing denominators, we have

Note: The values of A and B can be quickly found using the technique described in the margin of Example $6.5.3$.}

$$\begin{aligned} 19x+30 &= A(x+3) + B(x-5) && (6.5.17) \\ &= (A+B)x + (3A-5B). && (6.5.18) \end{aligned}$$

This implies that:

$$19 = A+B \tag{6.5.19}$$

$$30 = 3A - 5B. \tag{6.5.20}$$

Solving this system of linear equations gives

$$125/8 = A \tag{6.5.21}$$
$$27/8 = B. \tag{6.5.22}$$

We can now integrate.

$$\int \frac{x^3}{(x-5)(x+3)}\ dx = \int\left(x+2+\frac{125/8}{x-5}+\frac{27/8}{x+3}\right)\ dx \tag{6.5.23}$$
$$= \frac{x^2}{2}+2x+\frac{125}{8}\ln|x-5|+\frac{27}{8}\ln|x+3|+C. \tag{6.5.24}$$

Example 6.5.5: Integrating using partial fractions

Use partial fraction decomposition to evaluate $\int\frac{7x^2+31x+54}{(x+1)(x^2+6x+11)}\ dx$.

Solution

The degree of the numerator is less than the degree of the denominator so we begin by applying Key Idea 15. We have:

$$\frac{7x^2+31x+54}{(x+1)(x^2+6x+11)} = \frac{A}{x+1}+\frac{Bx+C}{x^2+6x+11}. \tag{6.5.25}$$

Now clear the denominators.

$$7x^2+31x+54 = A(x^2+6x+11)+(Bx+C)(x+1) \tag{6.5.26}$$
$$= (A+B)x^2+(6A+B+C)x+(11A+C). \tag{6.5.27}$$

This implies that:

$$7 = A+B \tag{6.5.28}$$
$$31 = 6A+B+C \tag{6.5.29}$$
$$54 = 11A+C. \tag{6.5.30}$$

Solving this system of linear equations gives the nice result of $A=5$, $B=2$ and $C=-1$. Thus

$$\int\frac{7x^2+31x+54}{(x+1)(x^2+6x+11)}\ dx = \int\left(\frac{5}{x+1} + \frac{2x-1}{x^2+6x+11}\right)\ dx.\]

The first term of this new integrand is easy to evaluate; it leads to a $5\ln|x+1|$ term. The second term is not hard, but takes several steps and uses substitution techniques.

The integrand $\frac{2x-1}{x^2+6x+11}$ has a quadratic in the denominator and a linear term in the numerator. This leads us to try substitution. Let $u=x^2+6x+11$, so $du=(2x+6)\ dx$. The numerator is $2x-1$, not $2x+6$, but we can get a $2x+6$ term in the numerator by adding 0 in the form of "$7-7$."

$$\frac{2x-1}{x^2+6x+11} = \frac{2x-1+7-7}{x^2+6x+11} \tag{6.5.31}$$
$$= \frac{2x+6}{x^2+6x+11}-\frac{7}{x^2+6x+11}. \tag{6.5.32}$$

We can now integrate the first term with substitution, leading to a $\ln|x^2+6x+11|$ term. The final term can be integrated using arctangent. First, complete the square in the denominator:

$$\frac{7}{x^2+6x+11} = \frac{7}{(x+3)^2+2}.\]

An antiderivative of the latter term can be found using Theorem 6.1.3 and substitution:

$$\int \frac{7}{x^2+6x+11}\ dx = \frac{7}{\sqrt{2}}\tan^{-1}\left(\frac{x+3}{\sqrt{2}}\right)+C.\]

Let's start at the beginning and put all of the steps together.

$$\int\frac{7x^2+31x+54}{(x+1)(x^2+6x+11)}\ dx = \int\left(\frac{5}{x+1}+\frac{2x-1}{x^2+6x+11}\right)\ dx \tag{6.5.33}$$
$$= \int\frac{5}{\quad}\ dx + \int\frac{2x+6}{\quad}\ dx \tag{6.5.34}$$

$$\int \quad x+1 \qquad \int \quad x^2+6x+11$$

$$\begin{aligned} &-\int \frac{7}{x^2+6x+11}\,dx \\ &= 5\ln|x+1| + \ln|x^2+6x+11| \\ &-\frac{7}{\sqrt{2}}\tan^{-1}\left(\frac{x+3}{\sqrt{2}}\right)+C. \end{aligned} \tag{6.5.35}$$

As with many other problems in calculus, it is important to remember that one is not expected to "see" the final answer immediately after seeing the problem. Rather, given the initial problem, we break it down into smaller problems that are easier to solve. The final answer is a combination of the answers of the smaller problems.

Partial Fraction Decomposition is an important tool when dealing with rational functions. Note that at its heart, it is a technique of algebra, not calculus, as we are rewriting a fraction in a new form. Regardless, it is very useful in the realm of calculus as it lets us evaluate a certain set of "complicated" integrals.

The next section introduces new functions, called the Hyperbolic Functions. They will allow us to make substitutions similar to those found when studying Trigonometric Substitution, allowing us to approach even more integration problems.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/
- Integrated by Justin Marshall.

This page titled 6.5: Partial Fraction Decomposition is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

6.6: Hyperbolic Functions

The **hyperbolic functions** are a set of functions that have many applications to mathematics, physics, and engineering. Among many other applications, they are used to describe the formation of satellite rings around planets, to describe the shape of a rope hanging from two points, and have application to the theory of special relativity. This section defines the hyperbolic functions and describes many of their properties, especially their usefulness to calculus.

These functions are sometimes referred to as the "hyperbolic trigonometric functions" as there are many, many connections between them and the standard trigonometric functions. Figure $6.6.1$ demonstrates one such connection. Just as cosine and sine are used to define points on the circle defined by $x^2+y^2=1$, the functions **hyperbolic cosine** and **hyperbolic sine** are used to define points on the hyperbola $x^2-y^2=1$.

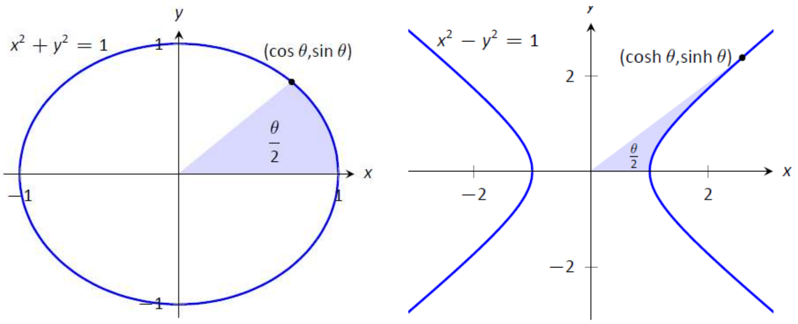


Figure $6.6.1$**:** Using trigonometric functions to define points on a circle and hyperbolic functions to define points on a hyperbola. The area of the shaded regions are included in them.

We begin with their definition.

Definition $6.6.1$: Hyperbolic Functions

1. $\cosh x=\frac{e^x+e^{-x}}{2}$
2. $\sinh x=\frac{e^x-e^{-x}}{2}$
3. $\tanh x=\frac{\sinh x}{\cosh x}$
4. $\text{sech} x=\frac{1}{\cosh x}$
5. $\text{csch} x=\frac{1}{\sinh x}$
6. $\coth x=\frac{\cosh x}{\sinh x}$

These hyperbolic functions are graphed in Figure $6.6.2$. In the graphs of $\cosh x$ and $\sinh x$, graphs of $e^x/2$ and $e^{-x}/2$ are included with dashed lines. As x gets "large," $\cosh x$ and $\sinh x$ each act like $e^x/2$; when x is a large negative number, $\cosh x$ acts like $e^{-x}/2$ whereas $\sinh x$ acts like $-e^{-x}/2$.

Notice the domains of $\tanh x$ and $\text{sech} x$ are $(-\infty,\infty)$, whereas both $\coth x$ and $\text{csch} x$ have vertical asymptotes at $x=0$. Also note the ranges of these functions, especially $\tanh x$: as $x\to\infty$, both $\sinh x$ and $\cosh x$ approach $e^{-x}/2$, hence $\tanh x$ approaches 1.

The following example explores some of the properties of these functions that bear remarkable resemblance to the properties of their trigonometric counterparts.

Pronunciation Note: "cosh" rhymes with "gosh," "sinh" rhymes with "pinch," and "tanh" rhymes with "ranch."

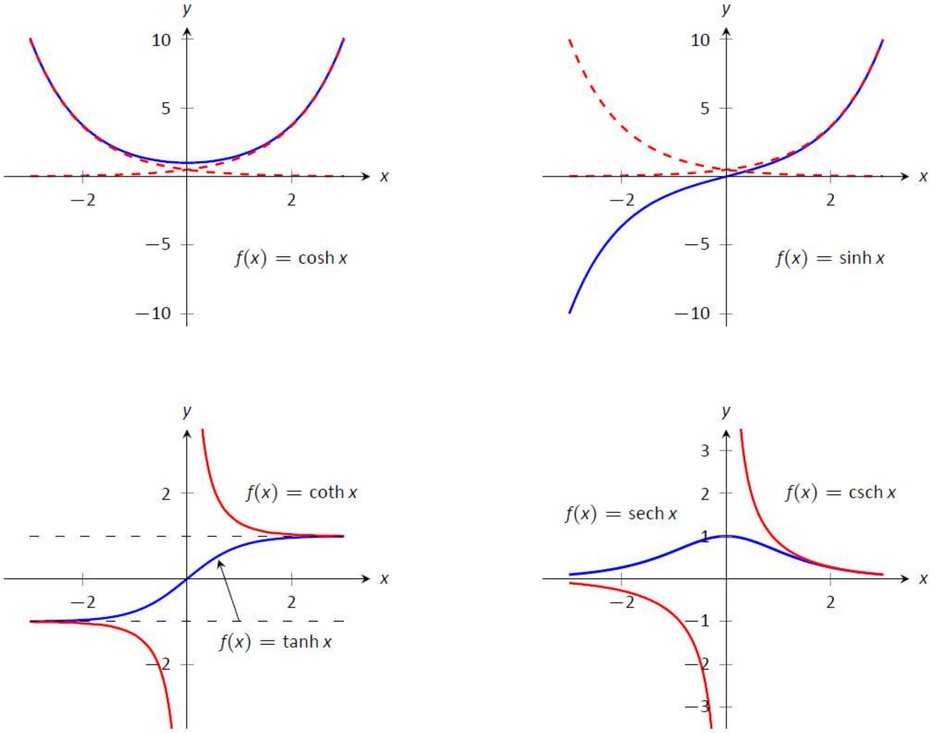


Figure $6.6.2$***:*** *Graphs of the hyperbolic functions.*

Example $6.6.1$: Exploring properties of hyperbolic functions

Use Definition $6.6.1$ to rewrite the following expressions.

1. $\cosh^2 x - \sinh^2 x$
2. $\tanh^2 x + \text{sech}^2 x$
3. $2\cosh x \sinh x$
4. $\frac{d}{dx}\big(\cosh x\big)$
5. $\frac{d}{dx}\big(\sinh x\big)$
6. $\frac{d}{dx}\big(\tanh x\big)$

Solution

1.
$$\begin{aligned}\cosh^2 x - \sinh^2 x &= \left(\frac{e^x+e^{-x}}{2}\right)^2 - \left(\frac{e^x-e^{-x}}{2}\right)^2 \qquad (6.6.1)\\ &= \frac{e^{2x}+2e^xe^{-x}+e^{-2x}}{4} - \frac{e^{2x}-2e^xe^{-x}+e^{-2x}}{4} \qquad (6.6.2)\\ &= \frac{4}{4} = 1. \qquad (6.6.3)\end{aligned}$$

So $\cosh^2 x - \sinh^2 x = 1$.

2.

Misplaced '#'

So $\tanh^2 x + \text{sech}^2 x = 1$.

3.
$$\begin{aligned}2\cosh x \sinh x &= 2\left(\frac{e^x+e^{-x}}{2}\right)\left(\frac{e^x-e^{-x}}{2}\right) \qquad (6.6.5)\\ &= 2\cdot\frac{e^{2x}-e^{-2x}}{4} \qquad (6.6.6)\\ &= \frac{e^{2x}-e^{-2x}}{2} = \sinh(2x). \qquad (6.6.7)\end{aligned}$$

Thus $2\cosh x \sinh x = \sinh(2x)$.

4.
$$\begin{aligned}\frac{d}{dx}\Big(\cosh x\Big) &= \frac{d}{dx}\left(\frac{e^x+e^{-x}}{2}\right) && (6.6.8)\\ &= \frac{e^x-e^{-x}}{2} && (6.6.9)\\ &= \sinh x. && (6.6.10)\end{aligned}$$

So $\frac{d}{dx}\Big(\cosh x\Big) = \sinh x.$

5.
$$\begin{aligned}\frac{d}{dx}\Big(\sinh x\Big) &= \frac{d}{dx}\left(\frac{e^x-e^{-x}}{2}\right) && (6.6.11)\\ &= \frac{e^x+e^{-x}}{2} && (6.6.12)\\ &= \cosh x. && (6.6.13)\end{aligned}$$

So $\frac{d}{dx}\Big(\sinh x\Big) = \cosh x.$

6.
$$\begin{aligned}\frac{d}{dx}\Big(\tanh x\Big) &= \frac{d}{dx}\left(\frac{\sinh x}{\cosh x}\right) && (6.6.14)\\ &= \frac{\cosh x\cosh x-\sinh x\sinh x}{\cosh^2 x} && (6.6.15)\\ &= \frac{1}{\cosh^2 x} && (6.6.16)\\ &= \text{sech}^2 x. && (6.6.17)\end{aligned}$$

So $\frac{d}{dx}\Big(\tanh x\Big) = \text{sech}^2 x.$

The following Key Idea summarizes many of the important identities relating to hyperbolic functions. Each can be verified by referring back to Definition 6.6.1.

Key Idea 16: Useful Hyperbolic Function Properties

Basic Identities

1. $\cosh^2 x-\sinh^2 x=1$
2. $\tanh^2 x+\text{sech}^2 x=1$
3. $\coth^2 x-\text{csch}^2 x=1$
4. $\cosh 2x=\cosh^2 x+\sinh^2 x$
5. $\sinh 2x\ =2\sinh x\cosh x$
6. $\cosh^2 x=\frac{\cosh 2x+1}{2}$
7. $\sinh^2 x=\frac{\cosh 2x-1}{2}$

Derivatives

1. $\frac{d}{dx}\Big(\cosh x\Big)\ =\sinh x$
2. $\frac{d}{dx}\Big(\sinh x\Big)\ =\cosh x$
3. $\frac{d}{dx}\Big(\tanh x\Big)\ =\text{sech}^2 x$
4. $\frac{d}{dx}\Big(\text{sech} x\Big)\ =-\text{sech} x\tanh x$
5. $\frac{d}{dx}\Big(\text{csch} x\Big)\ =-\text{csch} x\coth x$
6. $\frac{d}{dx}\Big(\coth x\Big)\ =-\text{csch}^2 x$

Integrals

1. $\int\cosh x\ dx=\sinh x+C$
2. $\int\sinh x\ dx=\cosh x+C$
3. $\int\tanh x\ dx=\ln(\cosh x)+C$
4. $\int\coth x\ dx=\ln|\sinh x|+C$

We practice using Key Idea 16

Example 6.6.2: Derivatives and integrals of hyperbolic functions

Evaluate the following derivatives and integrals.

1. $\frac{d}{dx}\left(\cosh 2x\right)$
2. $\int \operatorname{sech}^2(7t-3)\ dt$
3. $\int_0^{\ln 2} \cosh x\ dx$

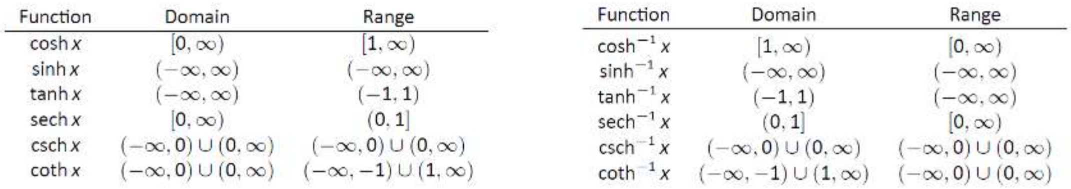
Solution

1. Using the Chain Rule directly, we have $\frac{d}{dx}\left(\cosh 2x\right) = 2\sinh 2x$.
   Just to demonstrate that it works, let's also use the Basic Identity found in Key Idea 16: $\cosh 2x = \cosh^2 x + \sinh^2 x$.
   $$\begin{aligned}\frac{d}{dx}\left(\cosh 2x\right) = \frac{d}{dx}\left(\cosh^2 x + \sinh^2 x\right) &= 2\cosh x \sinh x + 2\sinh x \cosh x \qquad (6.6.18)\\ &= 4\cosh x \sinh x. \qquad (6.6.19)\end{aligned}$$
   Using another Basic Identity, we can see that $4\cosh x \sinh x = 2\sinh 2x$. We get the same answer either way.
2. We employ substitution, with $u = 7t-3$ and $du = 7dt$. Applying Key Ideas 10 and 16 we have:
   $$\int \operatorname{sech}^2(7t-3)\ dt = \frac{1}{7}\tanh(7t-3) + C. \qquad (6.6.20)$$
3. $$\int_0^{\ln 2} \cosh x\ dx = \sinh x\Big|_0^{\ln 2} = \sinh(\ln 2) - \sinh 0 = \sinh(\ln 2). \qquad (6.6.21)$$
   We can simplify this last expression as $\sinh x$ is based on exponentials:
   $$\sinh(\ln 2) = \frac{e^{\ln 2} - e^{-\ln 2}}{2} = \frac{2 - 1/2}{2} = \frac{3}{4}. \qquad (6.6.22)$$

Inverse Hyperbolic Functions

Just as the inverse trigonometric functions are useful in certain integrations, the inverse hyperbolic functions are useful with others. Figure 16 shows the restrictions on the domains to make each function one-to-one and the resulting domains and ranges of their inverse functions. Their graphs are shown in Figure 6.6.3

Because the hyperbolic functions are defined in terms of exponential functions, their inverses can be expressed in terms of logarithms as shown in Key Idea 17. It is often more convenient to refer to $\sinh^{-1} x$ than to $\ln\left(x + \sqrt{x^2+1}\right)$, especially when one is working on theory and does not need to compute actual values. On the other hand, when computations are needed, technology is often helpful but many hand-held calculators lack a \textit{convenient} $\sinh^{-1} x$ button. (Often it can be accessed under a menu system, but not conveniently.) In such a situation, the logarithmic representation is useful. The reader is not encouraged to memorize these, but rather know they exist and know how to use them when needed.



Function	Domain	Range
$\cosh x$	$[0,\infty)$	$[1,\infty)$
$\sinh x$	$(-\infty,\infty)$	$(-\infty,\infty)$
$\tanh x$	$(-\infty,\infty)$	$(-1,1)$
$\operatorname{sech} x$	$[0,\infty)$	$(0,1]$
$\operatorname{csch} x$	$(-\infty,0)\cup(0,\infty)$	$(-\infty,0)\cup(0,\infty)$
$\coth x$	$(-\infty,0)\cup(0,\infty)$	$(-\infty,-1)\cup(1,\infty)$

Function	Domain	Range
$\cosh^{-1} x$	$[1,\infty)$	$[0,\infty)$
$\sinh^{-1} x$	$(-\infty,\infty)$	$(-\infty,\infty)$
$\tanh^{-1} x$	$(-1,1)$	$(-\infty,\infty)$
$\operatorname{sech}^{-1} x$	$(0,1]$	$[0,\infty)$
$\operatorname{csch}^{-1} x$	$(-\infty,0)\cup(0,\infty)$	$(-\infty,0)\cup(0,\infty)$
$\coth^{-1} x$	$(-\infty,-1)\cup(1,\infty)$	$(-\infty,0)\cup(0,\infty)$

Table* 6.6.1*: Graphs of* $\cosh x$*,* $\sinh x$ *and their inverses.

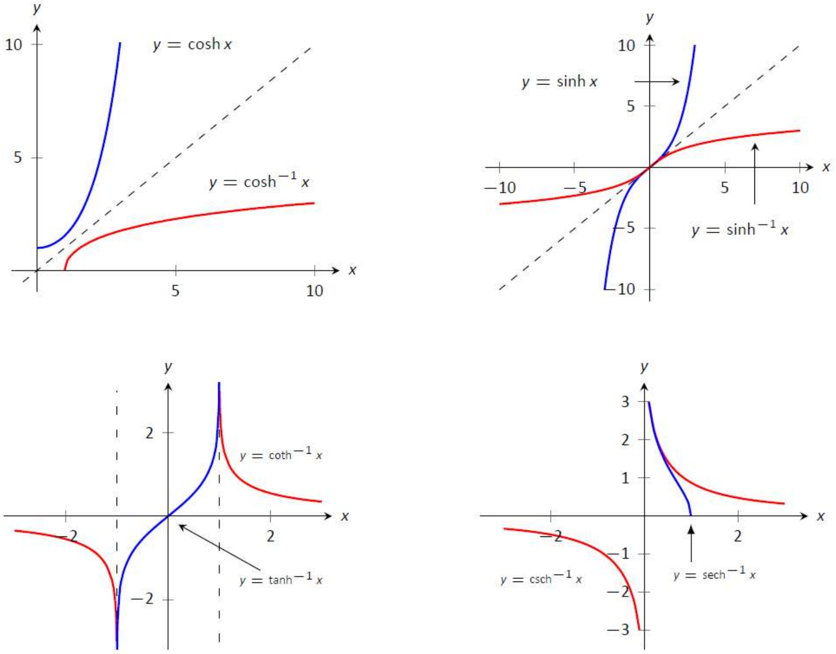


Figure 6.6.3**:** Graphs of the hyperbolic functions and their inverses.

The following Key Ideas give the derivatives and integrals relating to the inverse hyperbolic functions. In Key Idea 19, both the inverse hyperbolic and logarithmic function representations of the antiderivative are given, based on Key Idea 17. Again, these latter functions are often more useful than the former. Note how inverse hyperbolic functions can be used to solve integrals we used Trigonometric Substitution to solve in Section 6.4.

Key IDea 17: Logarithmic definitions of the inverse hyperbolic functions.

1. $\cosh^{-1} x = \ln\left(x + \sqrt{x^2-1}\right);\ x \geq 1$
2. $\tanh^{-1} x = \frac{1}{2}\ln\left(\frac{1+x}{1-x}\right);\ |x| < 1$
3. $\operatorname{sech}^{-1} x = \ln\left(\frac{1+\sqrt{1-x^2}}{x}\right);\ 0 < x \leq 1$
4. $\sinh^{-1} x = \ln\left(x + \sqrt{x^2+1}\right)$
5. $\coth^{-1} x = \frac{1}{2}\ln\left(\frac{x+1}{x-1}\right);\ |x| > 1$
6. $\operatorname{csch}^{-1} x = \ln\left(\frac{1}{x} + \frac{\sqrt{1+x^2}}{|x|}\right);\ x \neq 0$

Key Idea 18: Derivatives Involving Inverse Hyperbolic Functions

1. $\frac{d}{dx}\left(\cosh^{-1} x\right) = \frac{1}{\sqrt{x^2-1}};\ x > 1$
2. $\frac{d}{dx}\left(\sinh^{-1} x\right) = \frac{1}{\sqrt{x^2+1}}$
3. $\frac{d}{dx}\left(\tanh^{-1} x\right) = \frac{1}{1-x^2};\ |x| < 1$
4. $\frac{d}{dx}\left(\operatorname{sech}^{-1} x\right) = \frac{-1}{x\sqrt{1-x^2}};0 < x < 1$
5. $\frac{d}{dx}\left(\operatorname{csch}^{-1} x\right) = \frac{-1}{|x|\sqrt{1+x^2}};\ x \neq 0$
6. $\frac{d}{dx}\left(\coth^{-1} x\right) = \frac{1}{1-x^2};\ |x| > 1$

Key Idea 19: Integrals Involving Inverse Hyperbolic Functions

1. $\int \frac{1}{\sqrt{x^2-a^2}}\,dx = \cosh^{-1}\left(\frac{x}{a}\right) + C;\ 0 < a < x \quad = \ln\left|x + \sqrt{x^2-a^2}\right| + C$
2. $\int \frac{1}{\sqrt{x^2+a^2}}\,dx = \sinh^{-1}\left(\frac{x}{a}\right) + C;\ a > 0 \quad = \ln\left|x + \sqrt{x^2+a^2}\right| + C$
3. $\int \frac{1}{a^2-x^2}\,dx = \begin{cases} \frac{1}{a}\tanh^{-1}\left(\frac{x}{a}\right) + C & x^2 < a^2 \\ \frac{1}{a}\coth^{-1}\left(\frac{x}{a}\right) + C & a^2 < x^2 \end{cases} = \frac{1}{2a}\ln\left|\frac{a+x}{a-x}\right| + C$
4. $\int \frac{1}{x\sqrt{a^2-x^2}}\,dx = -\frac{1}{a}\operatorname{sech}^{-1}\left(\frac{x}{a}\right) + C;\ 0 < x < a \quad = \frac{1}{a}\ln\left(\frac{x}{a+\sqrt{a^2-x^2}}\right) + C$

5. $\int \frac{1}{x\sqrt{x^2+a^2}}\,dx =$ $-\frac{1}{a}\operatorname{csch}^{-1}\left|\frac{x}{a}\right| + C;\ x \neq 0,\ a > 0$ $= \frac{1}{a}\ln\left|\frac{x}{a+\sqrt{a^2+x^2}}\right| + C$

We practice using the derivative and integral formulas in the following example.

Example $6.6.3$: Derivatives and integrals involving inverse hyperbolic functions

Evaluate the following.

1. $\frac{d}{dx}\left[\cosh^{-1}\left(\frac{3x-2}{5}\right)\right]$
2. $\int \frac{1}{x^2-1}\,dx$
3. $\int \frac{1}{\sqrt{9x^2+10}}\,dx$

Solution

1. Applying Key Idea 18 with the Chain Rule gives:

$$\frac{d}{dx}\left[\cosh^{-1}\left(\frac{3x-2}{5}\right)\right] = \frac{1}{\sqrt{\left(\frac{3x-2}{5}\right)^2-1}}\cdot\frac{3}{5}. \tag{6.6.23}$$

2. Multiplying the numerator and denominator by (-1) gives: $\int \frac{1}{x^2-1}\,dx = \int \frac{-1}{1-x^2}\,dx$. The second integral can be solved with a direct application of item #3 from Key Idea 19, with $a = 1$. Thus

$$\int \frac{1}{x^2-1}\,dx = -\int \frac{1}{1-x^2}\,dx \tag{6.6.24}$$

$$= \begin{cases} -\tanh^{-1}(x) + C & x^2 < 1 \\ \\ -\coth^{-1}(x) + C & 1 < x^2 \end{cases} \tag{6.6.25}$$

$$= -\frac{1}{2}\ln\left|\frac{x+1}{x-1}\right| + C \tag{6.6.26}$$

$$= \frac{1}{2}\ln\left|\frac{x-1}{x+1}\right| + C. \tag{6.6.27}$$

We should note that this exact problem was solved at the beginning of Section 6.5. In that example the answer was given as $\frac{1}{2}\ln|x-1| - \frac{1}{2}\ln|x+1| + C$. Note that this is equivalent to the answer given in Equation $6.6.29$, as $\ln(a/b) = \ln a - \ln b$.

3. This requires a substitution, then item #2 of Key Idea 19 can be applied.

Let $u = 3x$, hence $du = 3dx$. We have

$$\int \frac{1}{\sqrt{9x^2+10}}\,dx = \frac{1}{3}\int \frac{1}{\sqrt{u^2+10}}\,du. \tag{6.6.28}$$

Note $a^2 = 10$, hence $a = \sqrt{10}$. Now apply the integral rule.

$$= \frac{1}{3}\sinh^{-1}\left(\frac{3x}{\sqrt{10}}\right) + C \tag{6.6.29}$$

$$= \frac{1}{3}\ln\left|3x + \sqrt{9x^2+10}\right| + C. \tag{6.6.30}$$

This section covers a lot of ground. New functions were introduced, along with some of their fundamental identities, their derivatives and antiderivatives, their inverses, and the derivatives and antiderivatives of these inverses. Four Key Ideas were presented, each including quite a bit of information.

Do not view this section as containing a source of information to be memorized, but rather as a reference for future problem solving. Key Idea 19 contains perhaps the most useful information. Know the integration forms it helps evaluate and understand how to use the inverse hyperbolic answer and the logarithmic answer.

The next section takes a brief break from demonstrating new integration techniques. It instead demonstrates a technique of evaluating limits that return indeterminate forms. This technique will be useful in Section 6.8, where limits will arise in the evaluation of certain definite integrals.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/
- Integrated by Justin Marshall.

This page titled 6.6: Hyperbolic Functions is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

6.7: L'Hopital's Rule

While this chapter is devoted to learning techniques of integration, this section is not about integration. Rather, it is concerned with a technique of evaluating certain limits that will be useful in the following section, where integration is once more discussed.

Our treatment of limits exposed us to "0/0", an indeterminate form. If $\lim_{x\to c} f(x)=0$ and $\lim_{x\to c} g(x)=0$, we do not conclude that $\lim_{x\to c} f(x)/g(x)$ is $0/0$; rather, we use $0/0$ as notation to describe the fact that both the numerator and denominator approach 0. The expression 0/0 has no numeric value; other work must be done to evaluate the limit.

Other indeterminate forms exist; they are: ∞/∞, $0\cdot\infty$, $\infty-\infty$, 0^0, 1^∞ and ∞^0. Just as "0/0" does not mean "divide 0 by 0," the expression "∞/∞" does not mean "divide infinity by infinity." Instead, it means "a quantity is growing without bound and is being divided by another quantity that is growing without bound." We cannot determine from such a statement what value, if any, results in the limit. Likewise, "$0\cdot\infty$" does not mean "multiply zero by infinity." Instead, it means "one quantity is shrinking to zero, and is being multiplied by a quantity that is growing without bound." We cannot determine from such a description what the result of such a limit will be.

This section introduces L'Hôpital's Rule, a method of resolving limits that produce the indeterminate forms 0/0 and ∞/∞. We'll also show how algebraic manipulation can be used to convert other indeterminate expressions into one of these two forms so that our new rule can be applied.

Theorem 6.7.1: L'Hôpital's Rule

Let $\lim_{x\to c} f(x)=0$ and $\lim_{x\to c} g(x)=0$, where f and g are differentiable functions on an open interval I containing c, and $g'(x)\neq 0$ on I except possibly at c. Then

$$\lim_{x\to c}\frac{f(x)}{g(x)}=\lim_{x\to c}\frac{f'(x)}{g'(x)}. \tag{6.7.1}$$

We demonstrate the use of L'Hôpital's Rule in the following examples; we will often use "LHR" as an abbreviation of "L'Hôpital's"

Example 6.7.1: Using L'Hôpital's Rule

Evaluate the following limits, using L'Hôpital's Rule as needed.

1. $\lim_{x\to 0}\frac{\sin x}{x}$
2. $\lim_{x\to 1}\frac{\sqrt{x+3}-2}{1-x}$
3. $\lim_{x\to 0}\frac{x^2}{1-\cos x}$
4. $\lim_{x\to 2}\frac{x^2+x-6}{x^2-3x+2}$

Solution

1. We proved this limit is 1 in Example ??? using the Squeeze Theorem. Here we use L'Hôpital's Rule to show its power.

$$\lim_{x\to 0}\frac{\sin x}{x}\overset{\text{by LHR}}{=}\lim_{x\to 0}\frac{\cos x}{1}=1. \tag{6.7.2}$$

2. $\lim_{x\to 1}\frac{\sqrt{x+3}-2}{1-x}\overset{\text{by LHR}}{=}\lim_{x\to 1}\frac{\frac{1}{2}(x+3)^{-1/2}}{-1}=-\frac{1}{4}$.
3. $\lim_{x\to 0}\frac{x^2}{1-\cos x}\overset{\text{by LHR}}{=}\lim_{x\to 0}\frac{2x}{\sin x}$.
   This latter limit also evaluates to the 0/0 indeterminate form. To evaluate it, we apply L'Hôpital's Rule again.

$$\lim_{x\to 0}\frac{2x}{\sin x}\overset{\text{by LHR}}{=}\frac{2}{\cos x}=2. \tag{6.7.3}$$

   Thus $\lim_{x\to 0}\frac{x^2}{1-\cos x}=2$.
4. We already know how to evaluate this limit; first factor the numerator and denominator. We then have:

$$\lim_{x\to 2}\frac{x^2+x-6}{x^2-3x+2}=\lim_{x\to 2}\frac{(x-2)(x+3)}{(x-2)(x-1)}=\lim_{x\to 2}\frac{x+3}{x-1}=5. \tag{6.7.4}$$

   We now show how to solve this using L'Hôpital's Rule.

$$\lim_{x\to 2}\frac{x^2+x-6}{x^2-3x+2}\overset{\text{by LHR}}{=}\lim_{x\to 2}\frac{2x+1}{2x-3}=5. \tag{6.7.5}$$

Note that at each step where L'Hôpital's Rule was applied, it was *needed*: the initial limit returned the indeterminate form of "$0/0$." If the initial limit returns, for example, 1/2, then L'Hôpital's Rule does not apply.

The following theorem extends our initial version of L'Hôpital's Rule in two ways. It allows the technique to be applied to the indeterminate form ∞/∞ and to limits where x approaches $\pm\infty$.

Theorem 6.7.2: L'Hôpital's Rule, Part 2

1. Let $\lim_{x\to a} f(x) = \pm\infty$ and $\lim_{x\to a} g(x) = \pm\infty$, where f and g are differentiable on an open interval I containing a. Then

$$\lim_{x\to a}\frac{f(x)}{g(x)} = \lim_{x\to a}\frac{f'(x)}{g'(x)}. \tag{6.7.6}$$

2. Let f and g be differentiable functions on the open interval (a,∞) for some value a, where $g'(x)\neq 0$ on (a,∞) and $\lim_{x\to\infty} f(x)/g(x)$ returns either 0/0 or ∞/∞. Then

$$\lim_{x\to\infty}\frac{f(x)}{g(x)} = \lim_{x\to\infty}\frac{f'(x)}{g'(x)}. \tag{6.7.7}$$

   A similar statement can be made for limits where x approaches $-\infty$.

Example 6.7.2: L'Hôpital's Rule with limits involving ∞

Evaluate the following limits.

$$1.\ \lim_{x\to\infty}\frac{3x^2-100x+2}{4x^2+5x-1000} \qquad 2.\ \lim_{x\to\infty}\frac{e^x}{x^3}. \tag{6.7.8}$$

Solution

1. We can evaluate this limit already using Theorem ???; the answer is 3/4. We apply L'Hôpital's Rule to demonstrate its applicability.

$$\lim_{x\to\infty}\frac{3x^2-100x+2}{4x^2+5x-1000} \overset{\text{by LHR}}{=} \lim_{x\to\infty}\frac{6x-100}{8x+5} \overset{\text{by LHR}}{=} \lim_{x\to\infty}\frac{6}{8} = \frac{3}{4}. \tag{6.7.9}$$

2.

$$\lim_{x\to\infty}\frac{e^x}{x^3} \overset{\text{by LHR}}{=} \lim_{x\to\infty}\frac{e^x}{3x^2} \overset{\text{by LHR}}{=} \lim_{x\to\infty}\frac{e^x}{6x} \overset{\text{by LHR}}{=} \lim_{x\to\infty}\frac{e^x}{6} = \infty. \tag{6.7.10}$$

   Recall that this means that the limit does not exist; as x approaches ∞, the expression e^x/x^3 grows without bound. We can infer from this that e^x grows "faster" than x^3; as x gets large, e^x is far larger than x^3. (This has important implications in computing when considering efficiency of algorithms.)

L'Hôpital's Rule can only be applied to ratios of functions. When faced with an indeterminate form such as $0\cdot\infty$ or $\infty-\infty$, we can sometimes apply algebra to rewrite the limit so that L'Hôpital's Rule can be applied. We demonstrate the general idea in the next example.

Example 6.7.3: Applying L'Hôpital's Rule to other indeterminate forms

Evaluate the following limits.

1. $\lim_{x\to 0^+} x\cdot e^{1/x}$
2. $\lim_{x\to 0^-} x\cdot e^{1/x}$
3. $\lim_{x\to\infty} \ln(x+1)-\ln x$
4. $\lim_{x\to\infty} x^2-e^x$

Solution

1. As $x\to 0^+$, $x\to 0$ and $e^{1/x}\to\infty$. Thus we have the indeterminate form $0\cdot\infty$. We rewrite the expression $x\cdot e^{1/x}$ as $\frac{e^{1/x}}{1/x}$; now, as $x\to 0^+$, we get the indeterminate form ∞/∞ to which L'Hôpital's Rule can be applied.

$$\lim_{x\to 0^+} x\cdot e^{1/x} = \lim_{x\to 0^+}\frac{e^{1/x}}{1/x} \overset{\text{by LHR}}{=} \lim_{x\to 0^+}\frac{(-1/x^2)e^{1/x}}{-1/x^2} = \lim_{x\to 0^+} e^{1/x} = \infty. \tag{6.7.11}$$

Interpretation: $e^{1/x}$ grows "faster" than x shrinks to zero, meaning their product grows without bound.

2. As $x \to 0^-$, $x \to 0$ and $e^{1/x} \to e^{-\infty} \to 0$. The the limit evaluates to $0 \cdot 0$ which is not an indeterminate form. We conclude then that

$$\lim_{x\to 0^-} x \cdot e^{1/x} = 0. \tag{6.7.12}$$

3. This limit initially evaluates to the indeterminate form $\infty - \infty$. By applying a logarithmic rule, we can rewrite the limit as

$$\lim_{x\to\infty} \ln(x+1) - \ln x = \lim_{x\to\infty} \ln\left(\frac{x+1}{x}\right). \tag{6.7.13}$$

As $x \to \infty$, the argument of the $\ln$ term approaches ∞/∞, to which we can apply L'Hôpital's Rule.

$$\lim_{x\to\infty} \frac{x+1}{x} \overset{\text{by LHR}}{=} \frac{1}{1} = 1. \tag{6.7.14}$$

Since $x \to \infty$ implies $\frac{x+1}{x} \to 1$, it follows that

$$x \to \infty \quad \text{implies} \quad \ln\left(\frac{x+1}{x}\right) \to \ln 1 = 0. \tag{6.7.15}$$

Thus

$$\lim_{x\to\infty} \ln(x+1) - \ln x = \lim_{x\to\infty} \ln\left(\frac{x+1}{x}\right) = 0. \tag{6.7.16}$$

Interpretation: since this limit evaluates to 0, it means that for large x, there is essentially no difference between $\ln(x+1)$ and $\ln x$; their difference is essentially 0.

4. The limit $\lim_{x\to\infty} x^2 - e^x$ initially returns the indeterminate form $\infty - \infty$. We can rewrite the expression by factoring out x^2; $x^2 - e^x = x^2\left(1 - \frac{e^x}{x^2}\right)$. We need to evaluate how e^x/x^2 behaves as $x \to \infty$:

$$\lim_{x\to\infty} \frac{e^x}{x^2} \overset{\text{by LHR}}{=} \lim_{x\to\infty} \frac{e^x}{2x} \overset{\text{by LHR}}{=} \lim_{x\to\infty} \frac{e^x}{2} = \infty. \tag{6.7.17}$$

Thus $\lim_{x\to\infty} x^2(1 - e^x/x^2)$ evaluates to $\infty \cdot (-\infty)$, which is not an indeterminate form; rather, $\infty \cdot (-\infty)$ evaluates to $-\infty$. We conclude that $\lim_{x\to\infty} x^2 - e^x = -\infty$.

Interpretation: as x gets large, the difference between x^2 and e^x grows very large.

Indeterminate Forms 0^0, 1^∞ and ∞^0

When faced with an indeterminate form that involves a power, it often helps to employ the natural logarithmic function. The following Key Idea expresses the concept, which is followed by an example that demonstrates its use.

Key Idea 20: Evaluating Limits Involving Indeterminate Forms 0^0, 1^∞ and ∞^0

If $\lim_{x\to c} \ln\big(f(x)\big) = L$, then $\lim_{x\to c} f(x) = \lim_{x\to c} e^{\ln(f(x))} = e^{\,L}$.

Example $6.7.4$: Using L'Hôpital's Rule with indeterminate forms involving exponents

Evaluate the following limits.

1. $\lim_{x\to\infty}\left(1 + \frac{1}{x}\right)^x$
2. $\lim_{x\to 0^+} x^x$.

Solution

1. This equivalent to a special limit given in Theorem ???; these limits have important applications within mathematics and finance. Note that the exponent approaches ∞ while the base approaches 1, leading to the indeterminate form 1^∞. Let $f(x) = (1 + 1/x)^x$; the problem asks to evaluate $\lim_{x\to\infty} f(x)$. Let's first evaluate $\lim_{x\to\infty} \ln\big(f(x)\big)$.

$$\lim_{x\to\infty} \ln\big(f(x)\big) = \lim_{x\to\infty} \ln\left(1 + \frac{1}{x}\right)^x \tag{6.7.18}$$

$$= \lim_{x\to\infty} x \ln\left(1 + \frac{1}{x}\right) \tag{6.7.19}$$

$$= \lim_{x\to\infty} \frac{\ln\left(1+\frac{1}{x}\right)}{1/x} \tag{6.7.20}$$

This produces the indeterminate form 0/0, so we apply L'Hôpital's Rule.

$$= \lim_{x\to\infty} \frac{\frac{1}{1+1/x}\cdot(-1/x^2)}{(-1/x^2)} \tag{6.7.21}$$

$$= \lim_{x\to\infty} \frac{1}{1+1/x} \tag{6.7.22}$$

$$= 1. \tag{6.7.23}$$

Thus $\lim_{x\to\infty} \ln\big(f(x)\big) = 1$. We return to the original limit and apply Key Idea 20.

$$\lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x = \lim_{x\to\infty} f(x) = \lim_{x\to\infty} e^{\ln(f(x))} = e^1 = e. \tag{6.7.24}$$

2. This limit leads to the indeterminate form 0^0. Let $f(x) = x^x$ and consider first $\lim_{x\to 0^+}\ln\big(f(x)\big)$.

$$\lim_{x\to 0^+}\ln\big(f(x)\big) = \lim_{x\to 0^+}\ln\left(x^x\right) \tag{6.7.25}$$

$$= \lim_{x\to 0^+} x\ln x \tag{6.7.26}$$

$$= \lim_{x\to 0^+}\frac{\ln x}{1/x}. \tag{6.7.27}$$

This produces the indeterminate form $-\infty/\infty$ so we apply L'Hôpital's Rule.

$$= \lim_{x\to 0^+}\frac{1/x}{-1/x^2} \tag{6.7.28}$$

$$= \lim_{x\to 0^+} -x \tag{6.7.29}$$

$$= 0. \tag{6.7.30}$$

Thus $\lim_{x\to 0^+}\ln\big(f(x)\big) = 0$. We return to the original limit and apply Key Idea 20.

$$\lim_{x\to 0^+} x^x = \lim_{x\to 0^+} f(x) = \lim_{x\to 0^+} e^{\ln(f(x))} = e^0 = 1. \tag{6.7.31}$$

This result is supported by the graph of $f(x) = x^x$ given in Figure $6.7.1$.

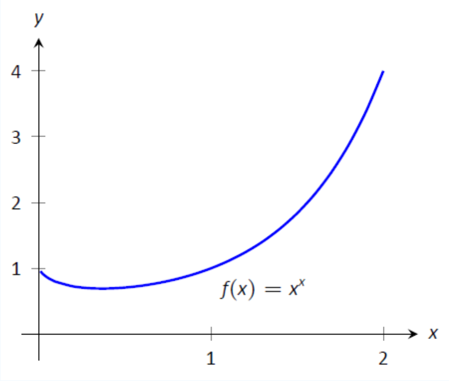


***Figure** $6.7.1$: A graph of $f(x) = x^x$ supporting the fact that as $x\to 0^+$, $f(x)\to 1$.*

Our brief revisit of limits will be rewarded in the next section where we consider *improper integration*. So far, we have only considered definite integrals where the bounds are finite numbers, such as $\int_0^1 f(x)\,dx$. Improper integration considers integrals where one, or both, of the bounds are "infinity." Such integrals have many uses and applications, in addition to generating ideas that are enlightening.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution -

Noncommercial (BY-NC) License. http://www.apexcalculus.com/

- Integrated by Justin Marshall.

This page titled 6.7: L'Hopital's Rule is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

6.8: Improper Integration

We begin this section by considering the following definite integrals:

$$\int_0^{100} \frac{1}{1+x^2}\, dx \approx 1.5608, \tag{6.8.1}$$

$$\int_0^{1000} \frac{1}{1+x^2}\, dx \approx 1.5698, \tag{6.8.2}$$

$$\int_0^{10,000} \frac{1}{1+x^2}\, dx \approx 1.5707. \tag{6.8.3}$$

Notice how the integrand is $1/(1+x^2)$ in each integral (which is sketched in Figure $6.8.1$). As the upper bound gets larger, one would expect the "area under the curve" would also grow. While the definite integrals do increase in value as the upper bound grows, they are not increasing by much. In fact, consider:

$$

$$\int_0^b \frac{1}{1+x^2}\, dx = \tan^{-1} x\Big|_0^b \tag{6.8.4}$$

$$= \tan^{-1} b - \tan^{-1} 0 \tag{6.8.5}$$

$$= \tan^{-1} b. \tag{6.8.6}$$

\]

As $b \to \infty$, $\tan^{-1} b \to \pi/2$. Therefore it seems that as the upper bound b grows, the value of the definite integral $\int_0^b \frac{1}{1+x^2}\, dx$ approaches $\pi/2 \approx 1.5708$. This should strike the reader as being a bit amazing: even though the curve extends "to infinity," it has a finite amount of area underneath it.

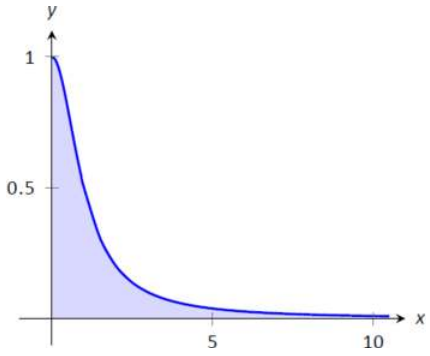


***Figure** $6.8.1$**: Graphing** $f(x) = \frac{1}{1+x^2}$*

When we defined the definite integral $\int_a^b f(x)\, dx$, we made two stipulations:

1. The interval over which we integrated, $[a, b]$, was a finite interval, and
2. The function $f(x)$ was continuous on $[a, b]$ (ensuring that the range of f was finite).

In this section we consider integrals where one or both of the above conditions do not hold. Such integrals are called **improper integrals**.

Improper Integrals with Infinite Bounds

Definition $6.8.1$: Improper Integrals with Infinite Bounds; Converge, Diverge

1. Let f be a continuous function on $[a, \infty)$. Define

$$\int_a^\infty f(x)\, dx \equiv \lim_{b\to\infty} \int_a^b f(x)\, dx. \tag{6.8.7}$$

2. Let f be a continuous function on $(-\infty, b]$. Define

$$\int_{-\infty}^{b} f(x)\ dx \equiv \lim_{a\to-\infty} \int_{a}^{b} f(x)\ dx. \tag{6.8.8}$$

3. Let f be a continuous function on $(-\infty, \infty)$. Let c be any real number; define

$$\int_{-\infty}^{\infty} f(x)\ dx \equiv \lim_{a\to-\infty} \int_{a}^{c} f(x)\ dx\ +\ \lim_{b\to\infty} \int_{c}^{b} f(x)\ dx. \tag{6.8.9}$$

An improper integral is said to **converge** if its corresponding limit exists; otherwise, it **diverges**. The improper integral in part 3 converges if and only if both of its limits exist.

Example 6.8.1: Evaluating improper integrals

Evaluate the following improper integrals.

1. $\int_1^\infty \frac{1}{x^2}\ dx$
2. $\int_1^\infty \frac{1}{x}\ dx$
3. $\int_{-\infty}^0 e^x\ dx$
4. $\int_{-\infty}^\infty \frac{1}{1+x^2}\ dx$

Solution

1.

$$[t] \int_1^\infty \frac{1}{x^2}\ dx\ =\ \lim_{b\to\infty} \int_1^b \frac{1}{x^2}\ dx\ =\ \lim_{b\to\infty} \frac{-1}{x}\bigg|_1^b \tag{6.8.10}$$

$$= \lim_{b\to\infty} \frac{-1}{b} + 1 \tag{6.8.11}$$

$$= 1. \tag{6.8.12}$$

A graph of the area defined by this integral is given in Figure 6.8.2.

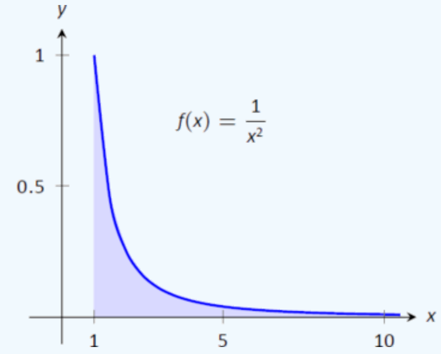


***Figure* 6.8.2*: A graph of* $f(x) = \frac{1}{x^2}$ *in Example* 6.8.1.**

2.

$$\int_1^\infty \frac{1}{x}\ dx = \lim_{b\to\infty} \int_1^b \frac{1}{x}\ dx \tag{6.8.13}$$

$$= \lim_{b\to\infty} \ln|x|\Big|_1^b \tag{6.8.14}$$

$$= \lim_{b\to\infty} \ln(b) \tag{6.8.15}$$

$$= \infty. \tag{6.8.16}$$

The limit does not exist, hence the improper integral $\int_1^\infty \frac{1}{x}\ dx$ diverges. Compare the graphs in Figures 6.8.3a and 6.8.3b; notice how the graph of $f(x) = 1/x$ is noticeably larger. This difference is enough to cause the improper integral to diverge.

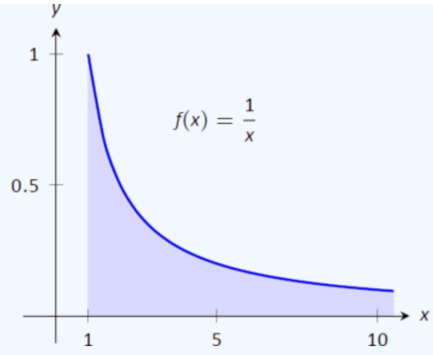


Figure $6.8.3$***:*** *A graph of* $f(x) = \frac{1}{x}$ *in Example* $6.8.1$

3. $$\int_{-\infty}^{0} e^x\ dx = \lim_{a\to-\infty} \int_a^0 e^x\ dx \tag{6.8.17}$$
$$= \lim_{a\to-\infty} e^x \Big|_a^0 \tag{6.8.18}$$
$$= \lim_{a\to-\infty} e^0 - e^a \tag{6.8.19}$$
$$= 1. \tag{6.8.20}$$

A graph of the area defined by this integral is given in Figure $6.8.4$.

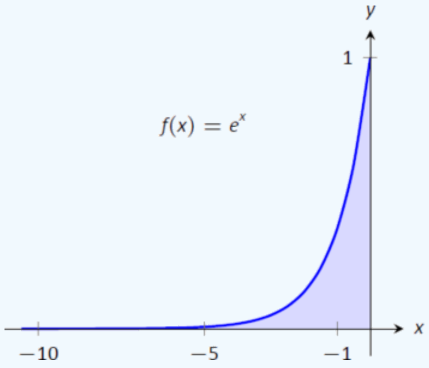


Figure $6.8.4$***:*** *A graph of* $f(x) = e^x$ *in Example* $6.8.1$

4. We will need to break this into two improper integrals and choose a value of c as in part 3 of Definition $6.8.1$. Any value of c is fine; we choose $c = 0$.

$$\int_{-\infty}^{\infty} \frac{1}{1+x^2}\ dx = \lim_{a\to-\infty} \int_a^0 \frac{1}{1+x^2}\ dx + \lim_{b\to\infty} \int_0^b \frac{1}{1+x^2}\ dx \tag{6.8.21}$$
$$= \lim_{a\to-\infty} \tan^{-1} x \Big|_a^0 + \lim_{b\to\infty} \tan^{-1} x \Big|_0^b \tag{6.8.22}$$
$$= \lim_{a\to-\infty} \left(\tan^{-1} 0 - \tan^{-1} a\right) + \lim_{b\to\infty} \left(\tan^{-1} b - \tan^{-1} 0\right) \tag{6.8.23}$$
$$= \left(0 - \frac{-\pi}{2}\right) + \left(\frac{\pi}{2} - 0\right). \tag{6.8.24}$$

Each limit exists, hence the original integral converges and has value:

$$= \pi. \tag{6.8.25}$$

A graph of the area defined by this integral is given in Figure $6.8.5$.

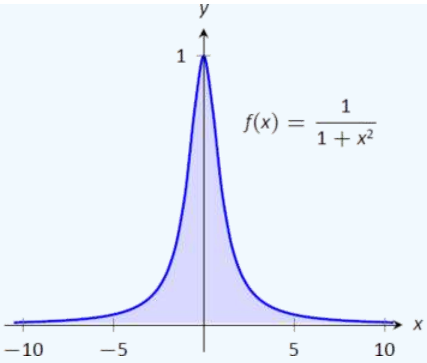


Figure 6.8.5**:** A graph of $f(x) = \frac{1}{1+x^2}$ in Example 6.8.1

The previous section introduced L'Hôpital's Rule, a method of evaluating limits that return indeterminate forms. It is not uncommon for the limits resulting from improper integrals to need this rule as demonstrated next.

Example 6.8.2: Improper integration and L'Hôpital's Rule

Evaluate the improper integral

$$\int_1^\infty \frac{\ln x}{x^2}\, dx. \tag{6.8.26}$$

Solution

This integral will require the use of Integration by Parts. Let $u = \ln x$ and $dv = 1/x^2\, dx$. Then

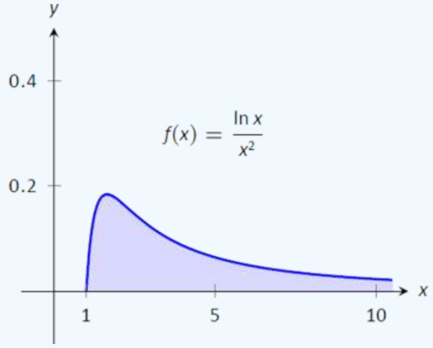


Figure *6.8.6**:** A graph of $f(x) = \frac{\ln x}{x^2}$ in Example 6.8.2*

$$\int_1^\infty \frac{\ln x}{x^2}\, dx = \lim_{b\to\infty} \int_1^b \frac{\ln x}{x^2}\, dx \tag{6.8.27}$$

$$= \lim_{b\to\infty} \left(-\frac{\ln x}{x}\bigg|_1^b + \int_1^b \frac{1}{x^2}\, dx \right) \tag{6.8.28}$$

$$= \lim_{b\to\infty} \left(-\frac{\ln x}{x} - \frac{1}{x} \right)\bigg|_1^b \tag{6.8.29}$$

$$= \lim_{b\to\infty} \left(-\frac{\ln b}{b} - \frac{1}{b} - (-\ln 1 - 1) \right). \tag{6.8.30}$$

The $1/b$ and $\ln 1$ terms go to 0, leaving $\lim_{b\to\infty} -\frac{\ln b}{b} + 1$. We need to evaluate $\lim_{b\to\infty} \frac{\ln b}{b}$ with L'Hôpital's Rule. We have:

$$\lim_{b\to\infty} \frac{\ln b}{b} \overset{\text{by LHR}}{=} \lim_{b\to\infty} \frac{1/b}{1} \tag{6.8.31}$$

$$= 0. \tag{6.8.32}$$

Thus the improper integral evaluates as:

$$\int_1^\infty \frac{\ln x}{x^2}\, dx = 1. \tag{6.8.33}$$

Improper Integrals with Infinite Range

We have just considered definite integrals where the interval of integration was infinite. We now consider another type of improper integration, where the range of the integrand is infinite.

Definition 6.8.2: Improper Integration with Infinite Range

{Let $f(x)$ be a continuous function on $[a,b]$ except at c, $a \le c \le b$, where $x=c$ is a vertical asymptote of f. Define

$$\int_a^b f(x)\ dx = \lim_{t\to c^-}\int_a^t f(x)\ dx + \lim_{t\to c^+}\int_t^b f(x)\ dx. \tag{6.8.34}$$

Example 6.8.3: Improper integration of functions with infinite range

Evaluate the following improper integrals:

1. $\int_0^1 \frac{1}{\sqrt{x}}\ dx$
2. $\int_{-1}^1 \frac{1}{x^2}\ dx.$

Solution

1. A graph of $f(x)=1/\sqrt{x}$ is given in Figure 6.8.7. Notice that f has a vertical asymptote at $x=0$; in some sense, we are trying to compute the area of a region that has no "top." Could this have a finite value?

$$\begin{align}\int_0^1 \frac{1}{\sqrt{x}}\ dx &= \lim_{a\to 0^+}\int_a^1 \frac{1}{\sqrt{x}}\ dx \tag{6.8.35}\\ &= \lim_{a\to 0^+} 2\sqrt{x}\Big|_a^1 \tag{6.8.36}\\ &= \lim_{a\to 0^+} 2\left(\sqrt{1}-\sqrt{a}\right) \tag{6.8.37}\\ &= 2. \tag{6.8.38}\end{align}$$

It turns out that the region does have a finite area even though it has no upper bound (strange things can occur in mathematics when considering the infinite).
Note: In Definition 6.8.1, c can be one of the endpoints (a or b). In that case, there is only one limit to consider as part of the definition.

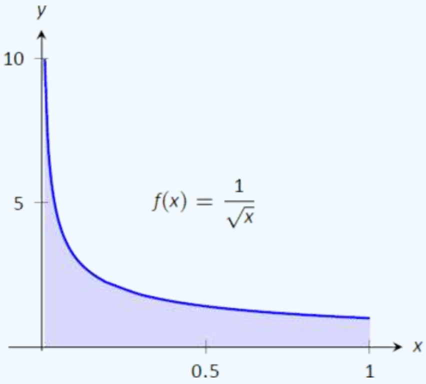


***Figure* 6.8.7*: A graph of* $f(x)=\frac{1}{\sqrt{x}}$ *in Example* 6.8.3**

2. The function $f(x)=1/x^2$ has a vertical asymptote at $x=0$, as shown in Figure 6.8.8, so this integral is an improper integral. Let's eschew using limits for a moment and proceed without recognizing the improper nature of the integral. This leads to:

$$\begin{align}\int_{-1}^1 \frac{1}{x^2}\ dx &= -\frac{1}{x}\Big|_{-1}^1 \tag{6.8.39}\\ &= -1-(1) \tag{6.8.40}\\ &= -2! \tag{6.8.41}\end{align}$$

Clearly the area in question is above the x-axis, yet the area is supposedly negative! Why does our answer not match our intuition? To answer this, evaluate the integral using Definition 6.8.2.

$$\int_{-1} \frac{1}{x^2}\,dx = \lim_{t\to 0^-}\int_{-1} \frac{1}{x^2}\,dx + \lim_{t\to 0^+}\int_{t} \frac{1}{x^2}\,dx \tag{6.8.42}$$

$$= \lim_{t\to 0^-} -\frac{1}{x}\Big|_{-1}^{t} + \lim_{t\to 0^+} -\frac{1}{x}\Big|_{t}^{1} \tag{6.8.43}$$

$$= \lim_{t\to 0^-} -\frac{1}{t} - 1 + \lim_{t\to 0^+} -1 + \frac{1}{t} \tag{6.8.44}$$

$$\Rightarrow \Big(\infty - 1\Big) + \Big(-1 + \infty\Big). \tag{6.8.45}$$

Neither limit converges hence the original improper integral diverges. The nonsensical answer we obtained by ignoring the improper nature of the integral is just that: nonsensical.

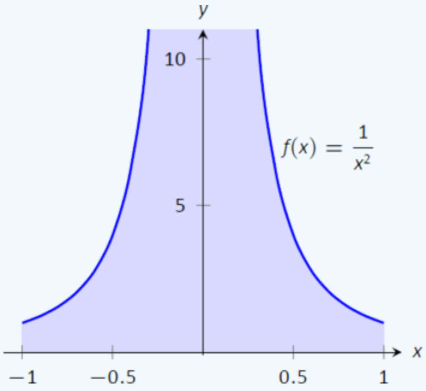


Figure $6.8.8$***:*** *A graph of* $f(x)=\frac{1}{x^2}$ *in Example* $6.8.3$

Understanding Convergence and Divergence

Oftentimes we are interested in knowing simply whether or not an improper integral converges, and not necessarily the value of a convergent integral. We provide here several tools that help determine the convergence or divergence of improper integrals without integrating.

Our first tool is to understand the behavior of functions of the form $\frac{1}{x^p}$.

Example $6.8.4$: Improper integration of $1/x^p$

Determine the values of p for which $\int_1^\infty \frac{1}{x^p}\,dx$ converges.

Solution

We begin by integrating and then evaluating the limit.

$$\int_1^\infty \frac{1}{x^p}\,dx = \lim_{b\to\infty}\int_1^b \frac{1}{x^p}\,dx \tag{6.8.46}$$

$$= \lim_{b\to\infty}\int_1^b x^{-p}\,dx \qquad (\text{assume } p\neq 1) \tag{6.8.47}$$

$$= \lim_{b\to\infty} \frac{1}{-p+1}x^{-p+1}\Big|_1^b \tag{6.8.48}$$

$$= \lim_{b\to\infty} \frac{1}{1-p}\big(b^{1-p} - 1^{1-p}\big). \tag{6.8.49}$$

When does this limit converge -- i.e., when is this limit *not* ∞? This limit converges precisely when the power of b is less than 0: when $1-p<0 \Rightarrow 1<p$.

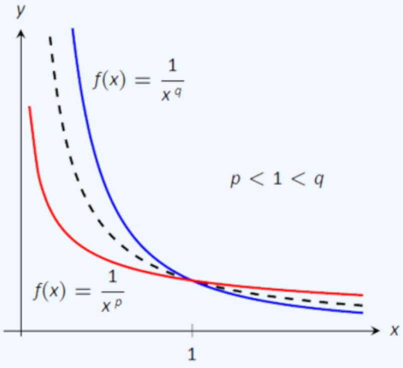


***Figure** $6.8.9$**:** Plotting functions of the form $1/x^p$ in Example $6.8.4$*

Our analysis shows that if $p > 1$, then $\int_1^\infty \frac{1}{x^p}\,dx$ converges. When $p < 1$ the improper integral diverges; we showed in Example $6.8.1$ that when $p = 1$ the integral also diverges.

Figure $6.8.9$ graphs $y = 1/x$ with a dashed line, along with graphs of $y = 1/x^p$, $p < 1$, and $y = 1/x^q$, $q > 1$. Somehow the dashed line forms a dividing line between convergence and divergence.

The result of Example $6.8.4$ provides an important tool in determining the convergence of other integrals. A similar result is proved in the exercises about improper integrals of the form $\int_0^1 \frac{1}{x^p}\,dx$. These results are summarized in the following Key Idea.

Key Idea 21: Convergence of Improper Integrals $\int_1^\infty \frac{1}{x^p}\,dx$ and $\int_0^1 \frac{1}{x^p}\,dx$.

1. The improper integral $\int_1^\infty \frac{1}{x^p}\,dx$ converges when $p > 1$ and diverges when $p \le 1$.
2. The improper integral $\int_0^1 \frac{1}{x^p}\,dx$ converges when $p < 1$ and diverges when $p \ge 1$.

A basic technique in determining convergence of improper integrals is to compare an integrand whose convergence is unknown to an integrand whose convergence is known. We often use integrands of the form $1/x^p$ to compare to as their convergence on certain intervals is known. This is described in the following theorem.

Theorem $6.8.1$: Direct Comparison Test for Improper Integrals

Let f and g be continuous on $[a, \infty)$ where $0 \le f(x) \le g(x)$ for all x in $[a, \infty)$.

1. If $\int_a^\infty g(x)\,dx$ converges, then $\int_a^\infty f(x)\,dx$ converges.
2. If $\int_a^\infty f(x)\,dx$ diverges, then $\int_a^\infty g(x)\,dx$ diverges.

Note: We used the upper and lower bound of "1" in Key Idea 21 for convenience. It can be replaced by any a where $a > 0$.

Example $6.8.5$: Determining convergence of improper integrals

Determine the convergence of the following improper integrals.

1. $\int_1^\infty e^{-x^2}\,dx$
2. $\int_3^\infty \frac{1}{\sqrt{x^2-x}}\,dx$

Solution

1. The function $f(x) = e^{-x^2}$ does not have an antiderivative expressible in terms of elementary functions, so we cannot integrate directly. It is comparable to $g(x) = 1/x^2$, and as demonstrated in Figure $6.8.10$, $e^{-x^2} < 1/x^2$ on $[1, \infty)$. We know from Key Idea 21 that $\int_1^\infty \frac{1}{x^2}\,dx$ converges, hence $\int_1^\infty e^{-x^2}\,dx$ also converges.

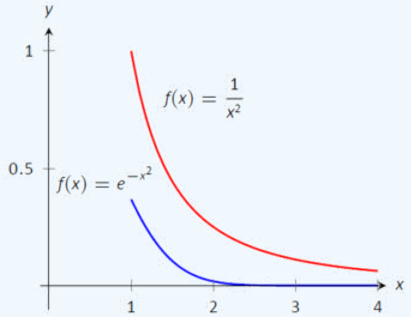


Figure 6.8.10***:*** *Graphs of* $f(x) = e^{-x^2}$ *and* $f(x) = 1/x^2$ *in Example* 6.8.6

2. Note that for large values of x, $\frac{1}{\sqrt{x^2-x}} \approx \frac{1}{\sqrt{x^2}} = \frac{1}{x}$. We know from Key Idea 21 and the subsequent note that $\int_3^\infty \frac{1}{x}\,dx$ diverges, so we seek to compare the original integrand to $1/x$.
   It is easy to see that when $x > 0$, we have $x = \sqrt{x^2} > \sqrt{x^2 - x}$. Taking reciprocals reverses the inequality, giving

$$\frac{1}{x} < \frac{1}{\sqrt{x^2-x}}. \tag{6.8.50}$$

   Using Theorem 6.8.1, we conclude that since $\int_3^\infty \frac{1}{x}\,dx$ diverges, $\int_3^\infty \frac{1}{\sqrt{x^2-x}}\,dx$ diverges as well. Figure 6.8.11 illustrates this.

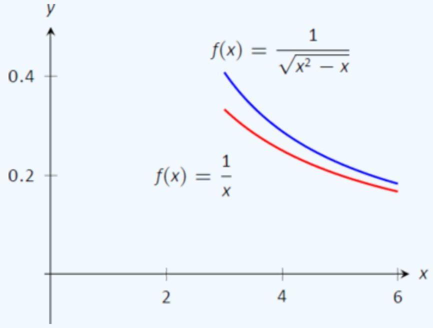


Figure 6.8.11***:*** *Graphs of* $f(x) = 1/\sqrt{x^2-x}$ *and* $f(x) = 1/x$ *in Example* 6.8.5

Being able to compare "unknown" integrals to "known" integrals is very useful in determining convergence. However, some of our examples were a little "too nice." For instance, it was convenient that $\frac{1}{x} < \frac{1}{\sqrt{x^2-x}}$, but what if the "$-x$" were replaced with a "$+2x + 5$"? That is, what can we say about the convergence of $\int_3^\infty \frac{1}{\sqrt{x^2+2x+5}}\,dx$? We have $\frac{1}{x} > \frac{1}{\sqrt{x^2+2x+5}}$, so we cannot use Theorem 6.8.1.

In cases like this (and many more) it is useful to employ the following theorem.

Theorem: Limit Comparison Test for Improper Integrals

Let f and g be continuous functions on $[a, \infty)$ where $f(x) > 0$ and $g(x) > 0$ for all x. If

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = L, \qquad 0 < L < \infty, \tag{6.8.51}$$

then

$$\int_a^\infty f(x)\,dx \quad \text{and} \quad \int_a^\infty g(x)\,dx \tag{6.8.52}$$

either both converge or both diverge.

Example 6.8.6: Determining convergence of improper integrals

Determine the convergence of $\int_3^\infty \frac{1}{\sqrt{x^2+2x+5}}\,dx$.

Solution

As x gets large, the quadratic inside the square root function will begin to behave much like $y=x$. So we compare $\frac{1}{\sqrt{x^2+2x+5}}$\ to $\frac{1}{x}$ with the Limit Comparison Test:

$$\lim_{x\to\infty} \frac{1/\sqrt{x^2+2x+5}}{1/x} = \lim_{x\to\infty}\frac{x}{\sqrt{x^2+2x+5}}.\]

The immediate evaluation of this limit returns ∞/∞, an indeterminate form. Using L'Hôpital's Rule seems appropriate, but in this situation, it does not lead to useful results. (We encourage the reader to employ L'Hôpital's Rule at least once to verify this.)

The trouble is the square root function. To get rid of it, we employ the following fact: If $\lim_{x\to c} f(x) = L$, then $\lim_{x\to c} f(x)^2 = L^2$. (This is true when either c or L is ∞.) So we consider now the limit\)

$$\lim_{x\to\infty} \frac{x^2}{x^2+2x+5}.\]

This converges to 1, meaning the original limit also converged to 1. As x gets very large, the function $\frac{1}{\sqrt{x^2+2x+5}}$ looks very much like $\frac{1}{x}$. Since we know that $\int_3^\infty \frac{1}{x}\,dx$ diverges, by the Limit Comparison Test we know that $\int_3^\infty \frac{1}{\sqrt{x^2+2x+5}}\,dx$ also diverges. Figure 6.8.12 graphs $f(x) = 1/\sqrt{x^2+2x+5}$ and $f(x) = 1/x$, illustrating that as x gets large, the functions become indistinguishable.

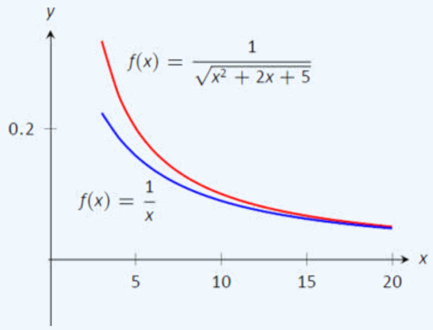


***Figure** 6.8.12: Graphing $f(x) = \frac{1}{\sqrt{x^2+2x+5}}$ and $f(x) = \frac{1}{x}$ in Example 6.8.6.*

Both the Direct and Limit Comparison Tests were given in terms of integrals over an infinite interval. There are versions that apply to improper integrals with an infinite range, but as they are a bit wordy and a little more difficult to employ, they are omitted from this text.

This chapter has explored many integration techniques. We learned Substitution, which "undoes" the Chain Rule of differentiation, as well as Integration by Parts, which "undoes" the Product Rule. We learned specialized techniques for handling trigonometric functions and introduced the hyperbolic functions, which are closely related to the trigonometric functions. All techniques effectively have this goal in common: rewrite the integrand in a new way so that the integration step is easier to see and implement.

As stated before, integration is, in general, hard. It is easy to write a function whose antiderivative is impossible to write in terms of elementary functions, and even when a function does have an antiderivative expressible by elementary functions, it may be really hard to discover what it is. The powerful computer algebra system *Mathematica* has approximately 1,000 pages of code dedicated to integration.

Do not let this difficulty discourage you. There is great value in learning integration techniques, as they allow one to manipulate an integral in ways that can illuminate a concept for greater understanding. There is also great value in understanding the need for good numerical techniques: the Trapezoidal and Simpson's Rules are just the beginning of powerful techniques for approximating the value of integration.

The next chapter stresses the uses of integration. We generally do not find antiderivatives for antiderivative's sake, but rather because they provide the solution to some type of problem. The following chapter introduces us to a number of different problems whose solution is provided by integration.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/
- Integrated by Justin Marshall.

This page titled 6.8: Improper Integration is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

6.E: Applications of Antidifferentiation (Exercises)

6.1: Substitution

Terms and Concepts

1. Substitution "undoes" what derivative Rule?

2. T/F: One can use algebra to rewrite the integrand of an integral to make it easier to evaluate.

Problems

In Exercises 3-14, evaluate the indefinite integrand to develop an understanding of Substitution.

3. $\int 3x^2(x^3-5)^7\,dx$

4. $\int (2x-5)(x^2-5x+7)^3\,dx$

5. $\int x(x^2+1)^8\,dx$

6. $\int (12x+14)(3x^2+7x+7)^3\,dx$

7. $\int \frac{1}{2x+7}\,dx$

8. $\int \frac{1}{\sqrt{2x+3}}\,dx$

9. $\int \frac{x}{\sqrt{x+3}}\,dx$

10. $\int \frac{x^3-x}{\sqrt{x}}\,dx$

11. $\int \frac{e^{\sqrt{x}}}{\sqrt{x}}\,dx$

12. $\int \frac{x^4}{\sqrt{x^5+1}}\,dx$

13. $\int \frac{\frac{1}{x}+1}{x^2}\,dx$

14. $\int \frac{\ln(x)}{x}\,dx$

In Exercises 15-23, use Substitution to evaluate the indefinite integral involving trigonometric functions.

15. $\int \sin^2(x)\cos(x)\,dx$

16. $\int \cos(3-6x)\,dx$

17. $\int \sec^2(4-x)\,dx$

18. $\int \sec(2x)\,dx$

19. $\int \tan^2(x)\sec^2(x)\,dx$

20. $\int x\cos(x^2)\,dx$

21. $\int \tan^2(x)\,dx$

22. $\int \cot x\,dx$. Do not just refer to Theorem 45 for the answer; justify it through Substitution.

23. $\int \csc x\,dx$. Do not just refer to Theorem 45 for the answer; justify it through Substitution.

In Exercises 24-30, use Substitution to evaluate the indefinite integral involving exponential functions.

24. $\int e^{3x-1}\,dx$

25. $\int e^{x^3}x^2\,dx$

26. $\int e^{x^2-2x+1}(x-1)\,dx$

27. $\int \frac{e^x+1}{e^x}\,dx$

28. $\int \frac{e^x-e^{-x}}{e^{2x}}\,dx$

29. $\int 3^{3x}\,dx$

30. $\int 4^{2x}\,dx$

In Exercises 31-34, use Substitution to evaluate the indefinite integral involving logarithmic functions.

31. $\int \frac{\ln x}{x}\, dx$

32. $\int \frac{(\ln x)^2}{x}\, dx$

33. $\int \frac{(\ln x)^3}{x}\, dx$

34. $\int \frac{1}{x\ln(x^2)}\, dx$

In Exercises 35-40, use Substitution to evaluate the indefinite integral involving rational functions.

35. $\int \frac{x^2+3x+1}{x}\, dx$

36. $\int \frac{x^3+x^2+x+1}{x}\, dx$

37. $\int \frac{x^3-1}{x+1}\, dx$

38. $\int \frac{x^2+2x-5}{x-3}\, dx$

39. $\int \frac{3x^2-5x+7}{x+1}\, dx$

40. $\int \frac{x^2+2x+1}{x^3+3x^2+3x}\, dx$

In Exercises 41-50, use Substitution to evaluate the indefinite integral inverse trigonometric functions.

41. $\int \frac{7}{x^2+7}\, dx$

42. $\int \frac{3}{\sqrt{9-x^2}}\, dx$

43. $\int \frac{14}{\sqrt{5-x^2}}\, dx$

44. $\int \frac{2}{x\sqrt{x^2-9}}\, dx$

45. $\int \frac{5}{\sqrt{x^4-16x^2}}\, dx$

46. $\int \frac{x}{\sqrt{1-x^4}}\, dx$

47. $\int \frac{1}{x^2-2x+8}\, dx$

48. $\int \frac{2}{\sqrt{-x^2+6x+7}}\, dx$

49. $\int \frac{3}{\sqrt{-x^2+8x+9}}\, dx$

50. $\int \frac{5}{x^2+6x+34}\, dx$

In Exercises 51-75, evaluate the indefinite integral.

51. $\int \frac{x^2}{(x^3+3)^2}\, dx$

52. $\int (3x^2+2x)(5x^3+5x^2+2)^8\, dx$

53. $\int \frac{x}{\sqrt{1-x^2}}\, dx$

54. $\int x^2\csc^2(x^3+1)\, dx$

55. $\int \sin(x)\sqrt{\cos(x)}\, dx$

56. $\int \frac{1}{x-5}\, dx$

57. $\int \frac{7}{3x+2}\, dx$

58. $\int \frac{3x^3+4x^2+2x-22}{x^2+3x+5}\, dx$

59. $\int \frac{2x+7}{x^2+7x+3}\, dx$

60. $\int \frac{9(2x+3)}{3x^2+9x+7}\, dx$

61. $\int \frac{-x^3+14x^2-46x-7}{x^2-7x+1}\, dx$

62. $\int \frac{x}{x^2+81}\, dx$

63. $\int \frac{2}{4x^2+1}\, dx$

64. $\int \frac{1}{x\sqrt{4x^2-1}}\,dx$

65. $\int \frac{1}{\sqrt{16-9x^2}}\,dx$

66. $\int \frac{3x-2}{x^2-2x+10}\,dx$

67. $\int \frac{7-2x}{x^2+12x+61}\,dx$

68. $\int \frac{x^2+5x-2}{x^2-10x+32}\,dx$

69. $\int \frac{x^3}{x^2+9}\,dx$

70. $\int \frac{x^3-x}{x^2+4x+9}\,dx$

71. $\int \frac{\sin(x)}{\cos^2(x)+1}\,dx$

72. $\int \frac{\cos(x)}{\sin^2(x)+1}\,dx$

73. $\int \frac{\cos(x)}{1-\sin^2(x)}\,dx$

74. $\int \frac{3x-3}{\sqrt{x^2-2x-6}}\,dx$

75. $\int \frac{x-3}{\sqrt{x^2-6x+8}}\,dx$

In Exercises 76-83, evaluate the definite integral.

76. $\int_1^3 \frac{1}{x-5}\,dx$

77. $\int_2^6 x\sqrt{x-2}\,dx$

78. $\int_{-\pi/2}^{\pi/2} \sin^2(x)\cos(x)\,dx$

79. $\int_0^1 2x(1-x^2)^4\,dx$

80. $\int_{-2}^{-1} (x+1)e^{x^2+2x+1}\,dx$

81. $\int_{-1}^{1} \frac{1}{x+x^2}\,dx$

82. $\int_2^4 \frac{1}{x^2-6x+10}\,dx$

83. $\int_1^{\sqrt{3}} \frac{1}{\sqrt{4-x^2}}\,dx$

6.2: Integration by Parts

Terms and Concepts

1. T/F: Integration by Parts is useful in evaluating integrands that contain products of function.

2. T/F: Integration by Parts can be thought of as the "opposite of the Chain Rule."

3. For what is "LIATE" useful?

Problems

In Exercises 4-33, evaluate the given indefinite integral.

4. $\int x\sin x\,dx$

5. $\int xe^{-x}\,dx$

6. $\int x^2\sin x\,dx$

7. $\int x^3\sin x\,dx$

8. $\int xe^{x^2}\,dx$

9. $\int x^3e^x\,dx$

10. $\int xe^{-2x}\,dx$

11. $\int e^x\sin x\,dx$

12. $\int e^{2x}\cos x\,dx$

13. $\int e^{2x}\sin(3x)\;dx$

14. $\int e^{5x}\cos(5x)\;dx$

15. $\int \sin x\cos x\,dx$

16. $\int \sin^{-1}x\,dx$

17. $\int \tan^{-1}(2x)\,dx$

18. $\int x\tan^{-1}x\,dx$

19. $\int \sin^{-1}x\,dx$

20. $\int x\ln x\,dx$

21. $\int (x-2)\ln x\;dx$

22. $\int x\ln(x-1)\;dx$

23. $\int x\ln(x^2)\,dx$

24. $\int x^2\ln x\,dx$

25. $\int (\ln x)^2\;dx$

26. $\int (\ln(x+1))^2\,dx$

27. $\int x\sec^2 x\,dx$

28. $\int x\csc^2 x\,dx$

29. $\int x\sqrt{x-2}\,dx$

30. $\int x\sqrt{x^2-2}\,dx$

31. $\int \sec x\tan x\,dx$

32. $\int x\sec x\tan x\,dx$

33. $\int x\csc x\cot x\;dx$

In Exercises 34-38, evaluate the indefinite integral after first making a substitution.

34. $\int \sin(\ln x)\;dx$

35. $\int \sin(\sqrt{x})\;dx$

36. $\int \ln(\sqrt{x})\;dx$

37. $\int e^{\sqrt{x}}\;dx$

38. $\int e^{\ln x}\;dx$

In Exercises 39-47, evaluate the definite integral. Note: the corresponding indefinite integrals appear in Exercises 4-12.

39. $\int_0^{\pi} x\sin x\,dx$

40. $\int_{-1}^{1} xe^{-x}\;dx$

41. $\int_{-\pi/4}{}^{\pi}/4x^2\sin x\,dx$

42. $\int_{-\pi/2}^{\pi/2} x^3\sin x\,dx$

43. $\int_0^{\sqrt{\ln 2}} xe^{x^2}\;dx$

44. $\int_0^{1} x^3e^{x}\;dx$

45. $\int_1^{2} xe^{-2x}\;dx$

46. $\int_0^{\pi} e^{x}\sin x\;dx$

47. $\int_{-\pi/2}^{\pi/2} e^{2x}\cos x\;dx$

6.3: Trigonometric Integrals

Terms and Concepts

1. T/F: $\int \sin^2(x)\cos^2 x\,dx$ cannot be evaluate using the techniques described in this section since both powers of $\sin x$ and $\cos x$ are even.

2. T/F: $\sin^3 x\cos^3 x\;dx$ cannot be evaluated using the techniques described in this section since both powers of $\sin x$ and $\cos x$ are odd.

3. T/F: This section addresses how to evaluate indefinite integrals such as $\int \sin^5 x\tan^3 x\,dx$.

Problems

In Exercises 4-26, evaluate the indefinite integral.

4. $\int \sin x\cos^4 x\,dx$

5. $\int \sin^3 x\cos x\,dx$

6. $\int \sin^3 x\cos^2 x\,dx$

7. $\int \sin^3 x\cos^3 x\,dx$

8. $\int \sin^6 x\cos^5 x\,dx$

9. $\int \sin^2 x\cos^7 x\,dx$

10. $\int \sin^2 x\cos^2 x\;dx$

11. $\int \sin(5x)\cos(3x)\;dx$

12. $\int \sin(x)\cos(2x)\;dx$

13. $\int \sin(3x)\sin(7x)\;dx$

14. $\int \sin(\pi x)\sin(2\pi x)\;dx$

15. $\int \cos(x)\cos(2x)\;dx$

16. $\int \cos\left(\frac{\pi}{2}x\right)\cos(\pi x)\;dx$

17. $\int \tan^4 x\sec^2 x\,dx$

18. $\int \tan^2 x\sec^4 x\,dx$

19. $\int \tan^3 x\sec^4 x\,dx$

20. $\int \tan^3 x\sec^2 x\,dx$

21. $\int \tan^3 x\sec^3 x\,dx$

22. $\int \tan^5 x\sec^5 x\,dx$

23. $\int \tan^4(x)\;dx$

24. $\int \sec^5 x\,dx$

25. $\int \tan^2 x\sec x\,dx$

26. $\int \tan^2 x\sec^3 x\,dx$

In Exercises 27-33, evaluate the definite integral. Note: the corresponding indefinite integrals appear in the previous set.

27. $\int_0^{\pi} \sin x\cos^4 x\,dx$

28. $\int_{-\pi}^{\pi} \sin^3 x\cos x\,dx$

29. $\int_{-\pi/2}^{\pi/2} \sin^2 x\cos^7 x\,dx$

30. $\int_0^{\pi/2} \sin(5x)\cos(3x)\,dx$

31. $\int_{-\pi/2}^{\pi/2} \cos(x)\cos(2x)\,dx$

32. $\int_0^{\pi/4} \tan^4 x \sec^2 x\, dx$

33. $\int_{-\pi/4}^{\pi/4} \tan^2 x \sec^4 x\, dx$

6.4: Trigonometric Substitution

Terms and Concepts

1. Trigonometric Substitution works on the same principles as Integration by Substitution, though it can feel "_____".

2. If one uses Trigonometric Substitution on an integrand containing $\sqrt{25-x^2}$, then one should set x = ______.

3. Consider the Pythagorean Identity $\sin^2\theta + \cos^2\theta = 1$.
(a) What identity is obtained when both sides are divided by $\cos^2\theta$?
(b) Use the new identity to simplify $9\tan^2\theta + 9$.

4. Why does Key Idea 13(a) state that $\sqrt{a^2-x^2} = a\cos\theta$, and not $|a\cos\theta|$?

Problems

In Exercises 5-16, apply Trigonometric Substitution to evaluate the indefinite integrals.

5. $\int \sqrt{x^2+1}\, dx$

6. $\int \sqrt{x^2+4}\, dx$

7. $\int \sqrt{1-x^2}\, dx$

8. $\int \sqrt{9-x^2}\, dx$

9. $\int \sqrt{x^2-1}\, dx$

10. $\int \sqrt{x^2-16}\, dx$

11. $\int \sqrt{4x^2+1}\, dx$

12. $\int \sqrt{1-9x^2}\, dx$

13. $\int \sqrt{16x^2-1}\, dx$

14. $\int \frac{3}{\sqrt{x^2+2}}\, dx$

15. $\int \frac{3}{\sqrt{7-x^2}}\, dx$

16. $\int \frac{5}{\sqrt{x^2-8}}\, dx$

In Exercises 17-26, evaluate the indefinite integrals. Some may be evaluated without Trigonometric Substitution.

17. $\int \frac{\sqrt{x^2-11}}{x}\, dx$

18. $\int \frac{1}{(x^2+1)^2}\, dx$

19. $\int \frac{x}{\sqrt{x^2-3}}\, dx$

20. $\int x^2\sqrt{1-x^2}\, dx$

21. $\int \frac{x}{(x^2+0)^{3/2}}\, dx$

22. $\int \frac{5x^2}{\sqrt{x^2-10}}\, dx$

23. $\int \frac{1}{(x^2+4x+13)^2}\, dx$

24. $\int x^2(1-x^2)^{-3/2}\, dx$

25. $\int \frac{\sqrt{5-x^2}}{7x^2}\, dx$

26. $\int \frac{x^2}{\sqrt{x^2+3}}\, dx$

In Exercises 27-32, evaluate the definite integrals by making the proper trigonometric substitution and changing the bounds of integration. (Note: each of the corresponding indefinite integrals has appeared previously in the Exercise set.)

27. $\int_{-1}^{1} \sqrt{1-x^2}\, dx$

28. $\int_4^8 \sqrt{x^2-16}\,dx$

29. $\int_0^2 \sqrt{x^2+4}\,dx$

30. $\int_{-1}^1 \frac{1}{(x^2+1)^2}\,dx$

31. $\int_{-1}^1 \sqrt{9x^2}\,dx$

32. $\int_{-1}^1 x^2\sqrt{1-x^2}\,dx$

6.5 Partial Fraction Decomposition

Terms and Concepts

1. Fill in the blank: Partial Fraction Decomposition is a method of rewriting _____ functions.

2. T/F: It is sometimes necessary to use polynomial division before using Partial Fraction Decomposition.

3. Decompose $\frac{1}{x^2-3x}$ without solving for the coefficients, as done in Example 181.

4. Decompose $\frac{7-x}{x^2-9}$ without solving for the coefficients, as done in Example 181.

5. Decompose $\frac{x-3}{x^2-7}$ without solving for the coefficients, as done in Example 181.

6. Decompose $\frac{2x+5}{x^3+7x}$ without solving for the coefficients, as done in Example 181.

Problems

In Exercises 7-25, evaluate the indefinite integral.

7. $\int \frac{7x+7}{x^2+3x-10}\,dx$

8. $\int \frac{7x-2}{x^2+x}\,dx$

9. $\int \frac{-4}{3x^2-12}\,dx$

10. $\int \frac{x+7}{(x+5)^2}\,dx$

11. $\int \frac{-3x-20}{(x+8)^2}\,dx$

12. $\int \frac{9x^2+11x+7}{x(x+1)^2}\,dx$

13. $\int \frac{-12x^2-x+33}{(x-1)(x+3)(3-2x)}\,dx$

14. $\int \frac{94x^2-10x}{(7x+3)(5x-1)(3x-1)}\,dx$

15. $\int \frac{x^2+2+1}{x^2+x-2}\,dx$

16. $\int \frac{x^3}{x^2-2x-20}\,dx$

17. $\int \frac{2x^2-4x+6}{x^2-2x+3}\,dx$

18. $\int \frac{1}{x^2+3x^2+3x}\,dx$

19. $\int \frac{x^2+x+5}{x^2+4x+10}\,dx$

20. $\int \frac{12x^2+21x+3}{(x+1)(3x^2+5x-1)}\,dx$

21. $\int \frac{6x^2+8x-4}{(x-3)(x^2+6x+10)}\,dx$

22. $\int \frac{2x^2+x+1}{(x+1)(x^2+9)}\,dx$

23. $\int \frac{x^2-20x-69}{(x-7)(x^2+2x+17)}\,dx$

24. $\int \frac{9x^2-60x+33}{(x-9)(x^2-2x+11)}\,dx$

25. $\int \frac{6x^2+45x+121}{(x+2)(x^2+10x+27)}\,dx$

In Exercises 26-29, evaluate the definite integral.

26. $\int_1^2 \frac{8x+21}{(x+2)(x+3)}\,dx$

27. $\int_0^5 \frac{14x+6}{(3x+2)(x+4)}\,dx$

28. $\int_{-1}^{1} \frac{x^2+5x-5}{(x-10)(x^2+4x+5)}\,dx$

29. $\int_{0}^{1} \frac{x}{(x+1)(x^2+2x+1)}\,dx$

6.6: Hyperbolic Functions

Terms and Concepts

1. In Key Idea 16, the equation $\int \tanh x\,dx = \ln(\cosh x) + C$ is given. Why is "$\ln|\cosh x|$" not used -i.e., why are absolute values no necessary?

2. The hyperbolic functions are used to define points on the right hand portion of the hyperbola $x^2 - y^2 = 1$, as shown in Figure 6.13. How can we use the hyperbolic functions to define points on the left hand portion of the hyperbola?

Problems

In Exercises 3-10, verify the given identity using Definition 23, as done in Example 186.

3. $\coth^2 x - \operatorname{csch}^2 x = 1$

4. $\cosh 2x = \cosh^2 x + \sinh^2 x$

5. $\cosh^2 x = \frac{\cosh 2x+1}{2}$

6. $\sinh^2 x = \frac{\cosh 2x-1}{2}$

7. $\frac{d}{dx}[\operatorname{sech} x] = -\operatorname{sech} x \tanh x$

8. $\frac{d}{dx}[\coth x] = -\operatorname{sech} x \tanh x$

9. $\int \tanh x\,dx = \ln(\cosh x) + C$

10. $\int \coth x\,dx = \ln|\sinh x| + C$

In Exercises 11-21, find the derivative of the given function.

11. $f(x) = \cosh 2x$

12. $f(x) = \tanh(x^2)$

13. $f(x) = \ln(\sinh x)$

14. $f(x) = \sinh x \cosh x$

15. $f(x) = x\sinh x - \cosh x$

16. $f(x) = \operatorname{sech}^{-1}(x^2)$

17. $f(x) = \sinh^{-1}(3x)$

18. $f(x) = \cosh^{-1}(2x^2)$

19. $f(x) = \tanh^{-1}(x+5)$

20. $f(x) = \tanh^{-1}(\cos x)$

21. $f(x) = \cosh^{-1}(\sec x)$

In Exercises 22-26, find the equation of the line tangent to the function at the given x-value.

22. $f(x) = \sinh x \text{ at } x = 0$

23. $f(x) = \cosh x \text{ at } x = \ln 2$

24. $f(x) = \operatorname{sech}^2 x \text{ at } x = \ln 3$

25. $f(x) = \sinh^{-1} x \text{ at } x = 0$

26. $f(x) = \cosh^{-1} x \text{ at } x = \sqrt{2}$

In Exercises 27-40, evaluate the given indefinite integral.

27. $\int \tanh(2x)\,dx$

28. $\int \cosh(3x-7)\,dx$

29. $\int \sinh x \cosh x\,dx$

30. $\int x\cosh x\,dx$

31. $\int x\sinh x\,dx$

32. $\int \frac{1}{9-x^2}\,dx$

33. $\int \frac{2x}{\sqrt{x^4-4}}\,dx$

34. $\int \frac{\sqrt{x}}{\sqrt{1+x^3}}\,dx$

35. $\int \frac{1}{x^2-16}\,dx$

36. $\int \frac{1}{x^2+x}\,dx$

37. $\int \frac{e^x}{x^{2x}+1}\,dx$

38. $\int \sinh^{-1} x\,dx$

39. $\int \tanh^{-1} x\,dx$

40. $\int \text{sech } x\,dx$ (Hint: multiply by $\frac{\cosh x}{\cosh x}$; set $u=\sinh x$.)

In Exercises 41-43, evaluate the given definite integral.

41. $\int_{-1}^{1} \sinh x\,dx$

42. $\int_{-\ln 2}^{\ln 2} \cosh x\,dx$

43. $\int_0^1 \tanh^{-1} x\,dx$.

6.7: L'Hopital's Rule

Terms and Concepts

1. List the different indeterminate forms described in this section.

2. T/F: l'Hopital's Rule provides a faster method of computing derivatives.

3. T/F: l'Hopitals Rule states that $\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right)=\frac{f'(x)}{g'(x)}$.

4. Explain what the indeterminate form "1^∞" means.

5. Fill in the blanks" The Quotient Rule is applied to $\frac{f(x)}{g(x)}$ when taking _____; l'Hopital's Rule is applied when taking certain_______.

6. Create (but do not evaluate) a limit that returns "∞^0".

7. Create a function $f(x)$ such that $\lim\limits_{x\to 1} f(x)$ returns "0^0".

Problems

In Exercises 8-52, evaluate the given limit.

8. $\lim\limits_{x\to 1} \frac{x^2+x-2}{x-1}$

9. $\lim\limits_{x\to 2} \frac{x^2+x-6}{x^2-7x+10}$

10. $\lim\limits_{x\to \pi} \frac{\sin x}{x-\pi}$

11. $\lim\limits_{x\to \pi/4} \frac{\sin x-\cos x}{\cos(2x)}$

12. $\lim\limits_{x\to 0} \frac{\sin(5x)}{x}$

13. $\lim\limits_{x\to 0} \frac{\sin(2x)}{x+2}$

14. $\lim\limits_{x\to 0} \frac{\sin(2x)}{\sin(3x)}$

15. $\lim\limits_{x\to 0} \frac{\sin(ax)}{\sin(bx)}$

16. $\lim\limits_{x\to 0^+} \frac{e^x-1}{x^2}$

17. $\lim\limits_{x\to 0^+} \frac{e^x-x-1}{x^2}$

18. $\lim\limits_{x\to 0^+} \frac{x-\sin x}{x^3-x^2}$

19. $\lim\limits_{x\to\infty} \frac{x^4}{e^x}$

20. $\lim\limits_{x\to\infty} \frac{\sqrt{x}}{e^x}$

21. $\lim\limits_{x\to\infty} \frac{e^x}{\sqrt{x}}$

22. $\lim\limits_{x\to\infty} \frac{e^x}{2^x}$

23. $\lim\limits_{x\to\infty} \frac{e^x}{3^x}$

24. $\lim\limits_{x\to 3} \frac{x^3-5x^2+3x+9}{x^3-7x^2+15x-9}$

25. $\lim\limits_{x\to -2} \frac{x^3+4x^2+4x}{x^3+7x^2+16x+12}$

26. $\lim\limits_{x\to\infty} \frac{\ln x}{x}$

27. $\lim\limits_{x\to\infty} \frac{\ln(x^2)}{x}$

28. $\lim\limits_{x\to\infty} \frac{(\ln x)^2}{x}$

29. $\lim\limits_{x\to 0^+} x\cdot\ln x$

30. $\lim\limits_{x\to 0^+} \sqrt{x}\cdot\ln x$

31. $\lim\limits_{x\to 0^+} xe^{1/x}$

32. $\lim\limits_{x\to\infty} x^3-x^2$

33. $\lim\limits_{x\to\infty} \sqrt{x}-\ln x$

34. $\lim\limits_{x\to -\infty} xe^x$

35. $\lim\limits_{x\to 0^+} \frac{1}{x^2}e^{-1/x}$

36. $\lim\limits_{x\to 0^+} (1+x)^{1/x}$

37. $\lim\limits_{x\to 0+} (2x)^x$

38. $\lim\limits_{x\to 0^+} (2/x)^x$

39. $\lim\limits_{x\to 0^+} (\sin x)^x$ Hint: use the Squeeze Theorem.

40. $\lim\limits_{x\to 1^+} (1-x)^{1-x}$

41. $\lim\limits_{x\to\infty} (x)^{1/x}$

42. $\lim\limits_{x\to\infty} (1/x)^x$

43. $\lim\limits_{x\to 1^1} (\ln x)^{1-x}$

44. $\lim\limits_{x\to\infty} (1+x)^{1/x}$

45. $\lim\limits_{x\to\infty} (1+x^2)^{1/x}$

46. $\lim\limits_{x\to\pi/2} \tan x\cos x$

47. $\lim\limits_{x\to\pi/2} \tan x\sin(2x)$

48. $\lim\limits_{x\to 1^+} \frac{1}{\ln x} - \frac{1}{1-x}$

49. $\lim\limits_{x\to 3^+} \frac{5}{x^2-9} - \frac{x}{x-3}$

50. $\lim\limits_{x\to\infty} x\tan(1/x)$

51. $\lim\limits_{x\to\infty} \frac{(\ln x)^3}{x}$

52. $\lim\limits_{x\to 1} \frac{x^2+x-2}{\ln x}$

6.8: Improper Integration

Terms and Concepts

1. The definite integral was defined with what two stipulations?

2. If $\lim\limits_{b\to\infty} \int_0^b f(x)\,dx$ exists, then the integral $\int_0^\infty f(x)\,dx$ is said to ___________.

3. If $\int_1^\infty f(x)\,dx = 10,\ \text{and}\ 0 \leq g(x) \leq f(x)$ for all *x*, then we know that $\int_1^\infty g(x)\,dx$ _______.

4. For what values of *p* will $\int_1^\infty \frac{1}{x^p}\,dx$ converge?

5. For what values of *p* will $\int_{10}^\infty \frac{1}{x^p}\,dx$ converge?

6. For what values of *p* will $\int_0^1 \frac{1}{x^p}\,dx$ converge?

Problems

In Exercises 7-33, evaluate the given improper integral.

7. $\int_0^\infty e^{5-2x}\,dx$

8. $\int_1^\infty \frac{1}{x^3}\,dx$

9. $\int_1^\infty x^{-4}\,dx$

10. $\int_{-\infty}^\infty \frac{1}{x^2+9}\,dx$

11. $\int_{-\infty}^0 2^x\,dx$

12. $\int_{-\infty}^0 \left(\frac{1}{2}\right)^x\,dx$

13. $\int_{-\infty}^\infty \frac{x}{x^2+1}\,dx$

14. $\int_{-\infty}^\infty \frac{x}{x^2+4}\,dx$

15. $\int_2^\infty \frac{1}{(x-1)^2}\,dx$

16. $\int_1^2 \frac{1}{(x-1)^2}\,dx$

17. $\int_2^\infty \frac{1}{x-1}\,dx$

18. $\int_1^2 \frac{1}{x-1}\,dx$

19. $\int_{-1}^1 \frac{1}{x}\,dx$

20. $\int_1^3 \frac{1}{x-2}\,dx$

21. $\int_0^\pi \sec^2 x\,dx$

22. $\int_{-2}^1 \frac{1}{\sqrt{|x|}}\,dx$

23. $\int_0^\infty xe^{-x}\,dx$

24. $\int_0^\infty xe^{-x^2}\,dx$

25. $\int_{-\infty}^\infty xe^{-x^2}\,dx$

26. $\int_{-\infty}^\infty \frac{1}{e^x+e^{-x}}\,dx$

27. $\int_0^1 x\ln x\,dx$

28. $\int_1^\infty \frac{\ln x}{x}\,dx$

29. $\int_0^1 \ln x \, dx$

30. $\int_1^\infty \frac{\ln x}{x^2} \, dx$

31. $\int_1^\infty \frac{\ln x}{\sqrt{x}} \, dx$

32. $\int_0^\infty e^{-x} \sin x \, dx$

33. $\int_0^\infty e^{-x} \cos x \, dx$

In Exercises 34-43, use the Direct Comparison Test or the Limit Comparison Test to determine whether the given definite integral converges or diverges. Clearly state what test is being used and what function the integrand is being compared to.

34. $\int_{10}^\infty \frac{3}{\sqrt{3x^2+2x-5}} \, dx$

35. $\int_2^\infty \frac{4}{\sqrt{7x^3-x}} \, dx$

36. $\int_0^\infty \frac{\sqrt{x+3}}{\sqrt{x^3-x^2+x+1}} \, dx$

37. $\int_1^\infty e^{-x} \ln x \, dx$

38. $\int_5^\infty e^{-x^2+3x-1} \, dx$

39. $\int_0^\infty \frac{\sqrt{x}}{e^x} \, dx$

40. $\int_2^\infty \frac{1}{x^2+\sin x} \, dx$

41. $\int_0^\infty \frac{x}{x^2+\cos x} \, dx$

42. $\int_0^\infty \frac{1}{x+e^x} \, dx$

43. $\int_0^\infty \frac{1}{e^x-x} \, dx$

6.E: Applications of Antidifferentiation (Exercises) is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

7: Applications of Integration

7.1: Area Between Curves
7.2: Volume by Cross-Sectional Area- Disk and Washer Methods
7.3: The Shell Method
7.4: Arc Length and Surface Area
7.5: Work
7.6: Fluid Forces
7.E: Applications of Integration (Exercises)

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 7: Applications of Integration is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

7.1: Area Between Curves

We begin this chapter with a reminder of a few key concepts from Chapter 5. Let f be a continuous function on $[a,b]$ which is partitioned into n subintervals as

$$a < x_1 < x_2 < \cdots < x_n < x_{n+1} = b. \tag{7.1.1}$$

Let dx_i denote the length of the i th subinterval, and let c_i be any x-value in that subinterval. Definition 5.3.1 states that the sum

$$\sum_{i=1}^{n} f(c_i)\ dx_i \tag{7.1.2}$$

is a *Riemann Sum.* Riemann Sums are often used to approximate some quantity (area, volume, work, pressure, etc.). The *approximation* becomes *exact* by taking the limit

$$\lim_{||dx_i|| \to 0} \sum_{i=1}^{n} f(c_i)\ dx_i, \tag{7.1.3}$$

where $||\ dx_i||$ the length of the largest subinterval in the partition. Theorem 5.3.2 connects limits of Riemann Sums to definite integrals:

$$\lim_{||dx_i|| \to 0} \sum_{i=1}^{n} f(c_i)\ dx_i = \int_a^b f(x)\ dx. \tag{7.1.4}$$

Finally, the Fundamental Theorem of Calculus states how definite integrals can be evaluated using antiderivatives.

This chapter employs the following technique to a variety of applications. Suppose the value Q of a quantity is to be calculated. We first approximate the value of Q using a Riemann Sum, then find the exact value via a definite integral. We spell out this technique in the following Key Idea.

Key Idea 22: Definite Integral Strategy

Let a quantity be given whose value Q is to be computed.

1. Divide the quantity into n smaller "subquantities" of value Q_i.
2. Identify a variable x and function $f(x)$ such that each subquantity can be approximated with the product $f(c_i)\ dx_i$, where dx_i represents a small change in x. Thus $Q_i \approx f(c_i)\ dx_i$. A sample approximation $f(c_i)\ dx_i$ of Q_i is called a *differential element*.
3. Recognize that $Q = \sum_{i=1}^{n} Q_i \approx \sum_{i=1}^{n} f(c_i)\ dx_i$, which is a Riemann Sum.
4. Taking the appropriate limit gives $Q = \int_a^b f(x)\ dx$

This Key Idea will make more sense after we have had a chance to use it several times. We begin with Area Between Curves, which we addressed briefly in Section 5.5.4.

Area Between Curves

We are often interested in knowing the area of a region. Forget momentarily that we addressed this already in Section 5.5.4 and approach it instead using the technique described in Key Idea 22.

Let Q be the area of a region bounded by continuous functions f and g. If we break the region into many subregions, we have an obvious equation:

$$\text{Total Area} = \text{sum of the areas of the subregions.} \tag{7.1.5}$$

The issue to address next is how to systematically break a region into subregions. A graph will help. Consider Figure $7.1.1a$ where a region between two curves is shaded. While there are many ways to break this into subregions, one particularly efficient way is to "slice" it vertically, as shown in Figure $7.1.1b$, into n equally spaced slices.

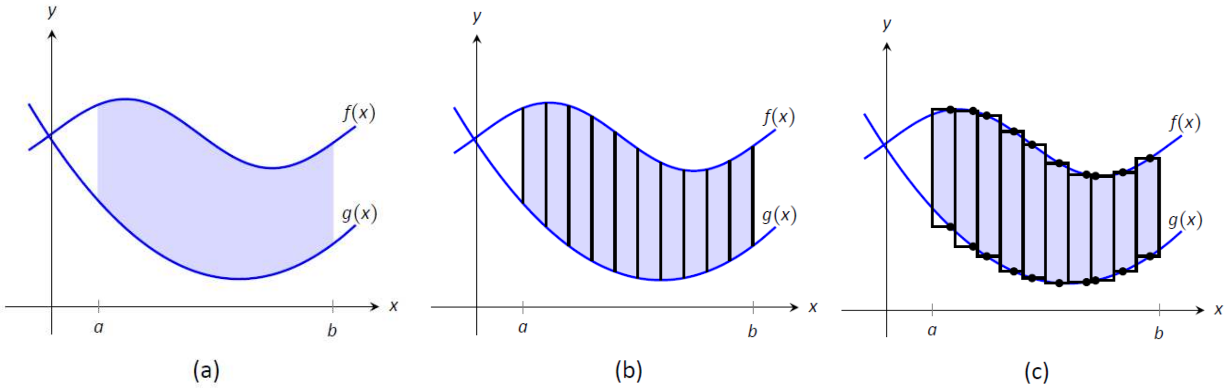


Figure 7.1.1***:*** *Subdividing a region into vertical slices and approximating the areas with rectangles.*

We now approximate the area of a slice. Again, we have many options, but using a rectangle seems simplest. Picking any x-value c_i in the $i^{\text{ th}}$ slice, we set the height of the rectangle to be $f(c_i)-g(c_i)$, the difference of the corresponding y-values. The width of the rectangle is a small difference in x-values, which we represent with dx. Figure 7.1.1c shows sample points c_i chosen in each subinterval and appropriate rectangles drawn. (Each of these rectangles represents a differential element.) Each slice has an area approximately equal to $\big(f(c_i)-g(c_i)\big)\ dx$; hence, the total area is approximately the Riemann Sum

$$Q=\sum_{i=1}^{n}\big(f(c_i)-g(c_i)\big)\ dx. \tag{7.1.6}$$

Taking the limit as $n\to\infty$ gives the exact area as $\int_a^b\big(f(x)-g(x)\big)\ dx$.

Theorem 7.1.1: Area Between Curves

Let $f(x)$ and $g(x)$ be continuous functions defined on $[a,b]$ where $f(x)\geq g(x)$ for all x in $[a,b]$. The area of the region bounded by the curves $y=f(x)$, $y=g(x)$ and the lines $x=a$ and $x=b$ is

$$\int_a^b\big(f(x)-g(x)\big)\ dx. \tag{7.1.7}$$

Example 7.1.1: Finding area enclosed by curves

Find the area of the region bounded by $f(x)=\sin x+2$, $g(x)=\frac{1}{2}\cos(2x)-1$, $x=0$ and $x=4\pi$, as shown in Figure 7.1.2.

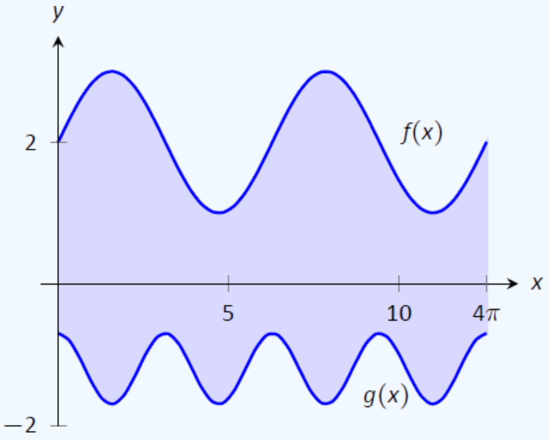


Figure 7.1.2***:*** *Graphing an enclosed region in Example* 7.1.1.

Solution

The graph verifies that the upper boundary of the region is given by f and the lower bound is given by g. Therefore the area of the region is the value of the integral

$$\begin{aligned}\int_0^{4\pi}\big(f(x)-g(x)\big)\,dx &= \int_0^{4\pi}\Big(\sin x+2-\big(\tfrac{1}{2}\cos(2x)-1\big)\Big)\,dx\\ &= -\cos x-\frac{1}{4}\sin(2x)+3x\Big|_0^{4\pi}\\ &= 12\pi\approx 37.7\text{ units}^2.\end{aligned}$$

Example 7.1.2: Finding total area enclosed by curves

Find the total area of the region enclosed by the functions $f(x)=-2x+5$ and $g(x)=x^3-7x^2+12x-3$ as shown in Figure 7.1.3.

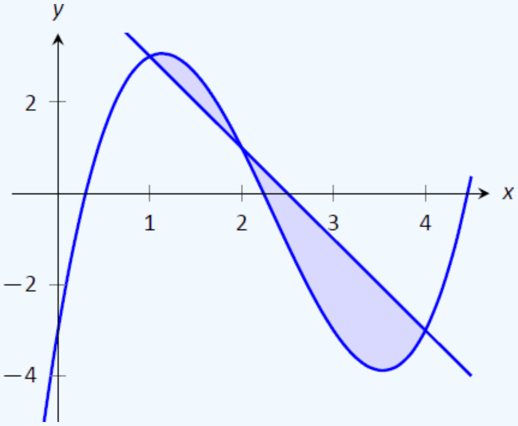


***FIgure** 7.1.3: Graphing a region enclosed by two functions in Example 7.1.2.*

Solution

A quick calculation shows that $f=g$ at $x=1,2$ and 4. One can proceed thoughtlessly by computing $\int_1^4\big(f(x)-g(x)\big)\,dx$, but this ignores the fact that on $[1,2]$, $g(x)>f(x)$. (In fact, the thoughtless integration returns $-9/4$, hardly the expected value of an *area*.) Thus we compute the total area by breaking the interval $[1,4]$ into two subintervals, $[1,2]$ and $[2,4]$ and using the proper integrand in each.

$$\begin{aligned}\text{Total Area} &= \int_1^2\big(g(x)-f(x)\big)\,dx+\int_2^4\big(f(x)-g(x)\big)\,dx\\ &= \int_1^2\big(x^3-7x^2+14x-8\big)\,dx+\int_2^4\big(-x^3+7x^2-14x+8\big)\,dx\\ &= 5/12+8/3\\ &= 37/12=3.083\text{ units}^2.\end{aligned}$$

The previous example makes note that we are expecting area to be *positive*. When first learning about the definite integral, we interpreted it as "signed area under the curve," allowing for "negative area." That doesn't apply here; area is to be positive.

The previous example also demonstrates that we often have to break a given region into subregions before applying Theorem 7.1.1. The following example shows another situation where this is applicable, along with an alternate view of applying the Theorem.

Example 7.1.3: Finding area: integrating with respect to y

Find the area of the region enclosed by the functions $y=\sqrt{x}+2$, $y=-(x-1)^2+3$ and $y=2$, as shown in Figure 7.1.4.

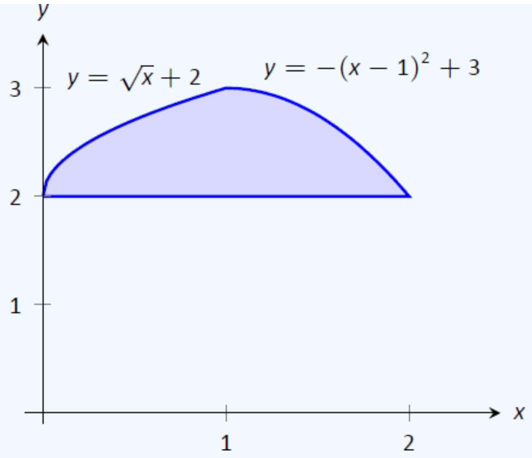


Figure 7.1.4*: Graphing a region for Example* 7.1.3.

Solution

We give two approaches to this problem. In the first approach, we notice that the region's "top" is defined by two different curves. On $[0,1]$, the top function is $y=\sqrt{x}+2$; on $[1,2]$, the top function is $y=-(x-1)^2+3$.

Thus we compute the area as the sum of two integrals:

$$\begin{aligned}\text{Total Area} &= \int_0^1 \left(\left(\sqrt{x}+2\right)-2\right)\,dx + \int_1^2 \left(\left(-(x-1)^2+3\right)-2\right)\,dx \\ &= 2/3 + 2/3 \\ &= 4/3.\end{aligned}$$

The second approach is clever and very useful in certain situations. We are used to viewing curves as functions of x; we input an x-value and a y-value is returned. Some curves can also be described as functions of y: input a y-value and an x-value is returned. We can rewrite the equations describing the boundary by solving for x:

$$y=\sqrt{x}+2 \quad \Rightarrow \quad x=(y-2)^2 \tag{7.1.8}$$

$$y=-(x-1)^2+3 \quad \Rightarrow \quad x=\sqrt{3-y}+1. \tag{7.1.9}$$

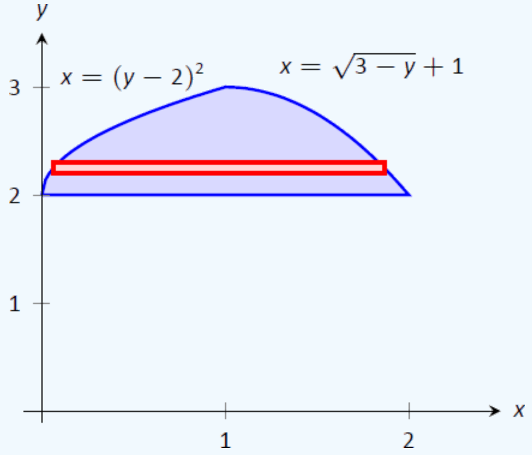


Figure 7.1.5*:The region used in Example* 7.1.3 *with boundaries relabeled as functions of* y.

Figure 7.1.5 shows the region with the boundaries relabeled. A differential element, a horizontal rectangle, is also pictured. The width of the rectangle is a small change in y: Δy. The height of the rectangle is a difference in x-values. The "top" x-value is the largest value, i.e., the rightmost. The "bottom" x-value is the smaller, i.e., the leftmost. Therefore the height of the rectangle is

$$\left(\sqrt{3-y}+1\right)-(y-2)^2. \tag{7.1.10}$$

The area is found by integrating the above function with respect to y with the appropriate bounds. We determine these by considering the y-values the region occupies. It is bounded below by $y=2$, and bounded above by $y=3$. That is, both the "top" and "bottom" functions exist on the y interval $[2,3]$. Thus

$$\begin{aligned}\text{Total Area} &= \int_2^3 \left(\sqrt{3-y}+1-(y-2)^2\right)\,dy\\ &= \left(-\frac{2}{3}(3-y)^{3/2}+y-\frac{1}{3}(y-2)^3\right)\Big|_2^3\\ &= 4/3.\end{aligned}$$

This calculus--based technique of finding area can be useful even with shapes that we normally think of as "easy." Example $7.1.4$ computes the area of a triangle. While the formula "$\frac{1}{2}\times\text{base}\times\text{height}$" is well known, in arbitrary triangles it can be nontrivial to compute the height. Calculus makes the problem simple.

Example $7.1.4$: Finding the area of a triangle

Compute the area of the regions bounded by the lines

$y=x+1$, $y=-2x+7$ and $y=-\frac{1}{2}x+\frac{5}{2}$, as shown in Figure $7.1.6$.

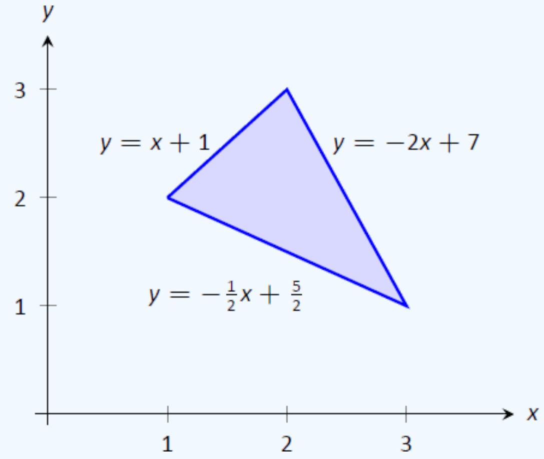


***Figure** $7.1.6$**:** Graphing a triangular region in Example $7.1.4$*

Solution

Recognize that there are two "top" functions to this region, causing us to use two definite integrals.

$$\begin{aligned}\text{Total Area} &= \int_1^2 \left((x+1)-(-\frac{1}{2}x+\frac{5}{2})\right)\,dx+\int_2^3 \left((-2x+7)-(-\frac{1}{2}x+\frac{5}{2})\right)\,dx\\ &= 3/4+3/4\\ &= 3/2.\end{aligned}$$

We can also approach this by converting each function into a function of y. This also requires 2 integrals, so there isn't really any advantage to doing so. We do it here for demonstration purposes.

The "top" function is always $x=\frac{7-y}{2}$ while there are two "bottom" functions. Being mindful of the proper integration bounds, we have

$$\begin{aligned}\text{Total Area} &= \int_1^2 \left(\frac{7-y}{2}-(5-2y)\right)\,dy+\int_2^3 \left(\frac{7-y}{2}-(y-1)\right)\,dy\\ &= 3/4+3/4\\ &= 3/2.\end{aligned}$$

Of course, the final answer is the same. (It is interesting to note that the area of all 4 subregions used is 3/4. This is coincidental.)

While we have focused on producing exact answers, we are also able to make approximations using the principle of Theorem 7.1.1. The integrand in the theorem is a distance ("top minus bottom"); integrating this distance function gives an area. By taking discrete measurements of distance, we can approximate an area using numerical integration techniques developed in Section ???. The following example demonstrates this.

Example 7.1.5: Numerically approximating area

To approximate the area of a lake, shown in Figure 7.1.7a, the "length" of the lake is measured at 200-foot increments as shown in Figure 7.1.7b, where the lengths are given in hundreds of feet. Approximate the area of the lake.

Solution

The measurements of length can be viewed as measuring "top minus bottom" of two functions. The exact answer is found by integrating $\int_0^{12}\big(f(x)-g(x)\big)\,dx$, but of course we don't know the functions f and g. Our discrete measurements instead allow us to approximate.

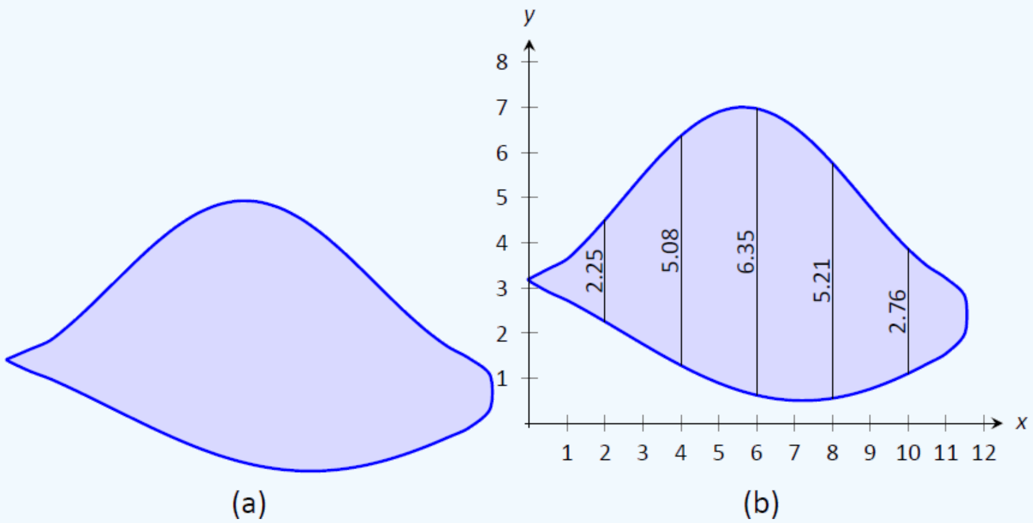


Figure 7.1.7***:*** *(a) A sketch of a lake, and (b) the lake with length measurements.*

We have the following data points:

$$(0,0),\ (2,2.25),\ (4,5.08),\ (6,6.35),\ (8,5.21),\ (10,2.76),\ (12,0). \tag{7.1.11}$$

We also have that $dx=\dfrac{b-a}{n}=2$, so Simpson's Rule gives

$$\begin{aligned}\text{Area} &\approx \frac{2}{3}\Big(1\cdot 0+4\cdot 2.25+2\cdot 5.08+4\cdot 6.35+2\cdot 5.21+4\cdot 2.76+1\cdot 0\Big)\\ &=44.01\overline{3}\ \text{units}^2.\end{aligned}$$

Since the measurements are in hundreds of feet, units $^2=(100\ \text{ft})^2=10,000\ \text{ft}^2$, giving a total area of $440,133\ \text{ft}^2$. (Since we are approximating, we'd likely say the area was about $440,000\ \text{ft}^2$, which is a little more than 10 acres.)

In the next section we apply our applications--of--integration techniques to finding the volumes of certain solids.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/
- Integrated by Justin Marshall.

This page titled 7.1: Area Between Curves is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

7.2: Volume by Cross-Sectional Area- Disk and Washer Methods

The volume of a general right cylinder, as shown in Figure 7.2.1, is

$$\text{Area of the base} \times \text{height}. \tag{7.2.1}$$

We can use this fact as the building block in finding volumes of a variety of shapes.

Given an arbitrary solid, we can *approximate* its volume by cutting it into n thin slices. When the slices are thin, each slice can be approximated well by a general right cylinder. Thus the volume of each slice is approximately its cross-sectional area × thickness. (These slices are the differential elements.)

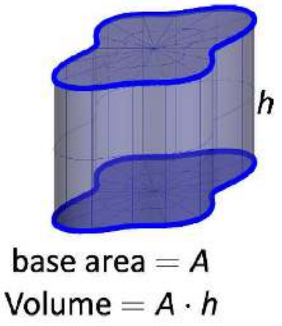


***Figure** 7.2.1**:** The volume of a general right cylinder.*

By orienting a solid along the x-axis, we can let $A(x_i)$ represent the cross-sectional area of the i^{th} slice, and let dx_i represent the thickness of this slice (the thickness is a small change in x). The total volume of the solid is approximately:

$$\text{Volume} \approx \sum_{i=1}^{n} \Big[\text{Area} \;\times\; \text{thickness}\Big] \tag{7.2.2}$$

$$= \sum_{i=1}^{n} A(x_i)\; dx_i. \tag{7.2.3}$$

Recognize that this is a Riemann Sum. By taking a limit (as the thickness of the slices goes to 0) we can find the volume exactly.

Theorem 7.2.1: Volume By Cross-Sectional Area

The volume V of a solid, oriented along the x-axis with cross-sectional area $A(x)$ from $x = a$ to $x = b$, is

$$V = \int_a^b A(x)\; dx. \tag{7.2.4}$$

Example 7.2.1: Finding the volume of a solid

Find the volume of a pyramid with a square base of side length 10 in and a height of 5 in.

Solution

There are many ways to "orient" the pyramid along the x-axis; Figure 7.2.2 gives one such way, with the pointed top of the pyramid at the origin and the x-axis going through the center of the base.

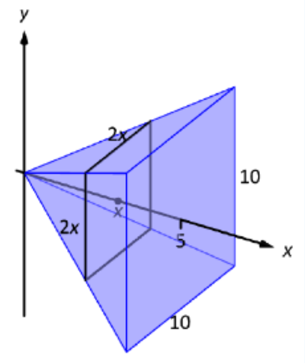


***Figure** 7.2.2: Orienting a pyramid along the x-axis in Example 7.2.1.*

Each cross section of the pyramid is a square; this is a sample differential element. To determine its area $A(x)$, we need to determine the side lengths of the square.

When $x=5$, the square has side length 10; when $x=0$, the square has side length 0. Since the edges of the pyramid are lines, it is easy to figure that each cross-sectional square has side length $2x$, giving $A(x)=(2x)^2=4x^2$.

If one were to cut a slice out of the pyramid at $x=3$, as shown in Figure 7.2.3, one would have a shape with square bottom and top with sloped sides. If the slice were thin, both the bottom and top squares would have sides lengths of about 6, and thus the cross--sectional area of the bottom and top would be about 36in². Letting Δx_i represent the thickness of the slice, the volume of this slice would then be about $36\Delta x_i\text{in}^3$.

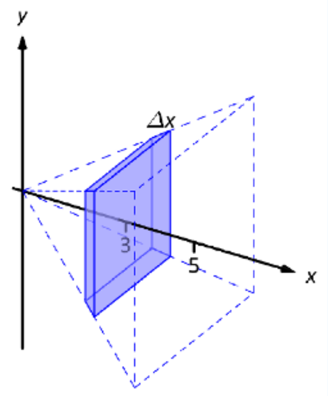


***Figure** 7.2.3: Cutting a slice in they pyramid in Example 7.2.1 at $x=3$.*

Cutting the pyramid into n slices divides the total volume into n equally--spaced smaller pieces, each with volume $(2x_i)^2\Delta x$, where x_i is the approximate location of the slice along the x-axis and Δx represents the thickness of each slice. One can approximate total volume of the pyramid by summing up the volumes of these slices:

$$\text{Approximate volume} = \sum_{i=1}^{n}(2x_i)^2\Delta x. \tag{7.2.5}$$

Taking the limit as $n\to\infty$ gives the actual volume of the pyramid; recoginizing this sum as a Riemann Sum allows us to find the exact answer using a definite integral, matching the definite integral given by Theorem 7.2.1.

We have

$$V = \lim_{n\to\infty}\sum_{i=1}^{n}(2x_i)^2\Delta x \tag{7.2.6}$$

$$= \int_0^5 4x^2\,dx \tag{7.2.7}$$

$$= \frac{4}{3}x^3\Big|_0^5 \tag{7.2.8}$$

$$= \frac{500}{3}\text{ in}^3 \approx 166.67\text{ in}^3. \tag{7.2.9}$$

We can check our work by consulting the general equation for the volume of a pyramid (see the back cover under "Volume of A General Cone"):

$$\frac{1}{3} \times \text{area of base} \times \text{height}. \tag{7.2.10}$$

Certainly, using this formula from geometry is faster than our new method, but the calculus--based method can be applied to much more than just cones.

An important special case of Theorem 7.2.1 is when the solid is a **solid of revolution**, that is, when the solid is formed by rotating a shape around an axis.

Start with a function $y = f(x)$ from $x = a$ to $x = b$. Revolving this curve about a horizontal axis creates a three-dimensional solid whose cross sections are disks (thin circles). Let $R(x)$ represent the radius of the cross-sectional disk at x; the area of this disk is $\pi R(x)^2$. Applying Theorem 7.2.1 gives the Disk Method.

Key Idea 23: The Disk Method

Let a solid be formed by revolving the curve $y = f(x)$ from $x = a$ to $x = b$ around a horizontal axis, and let $R(x)$ be the radius of the cross-sectional disk at x. The volume of the solid is

$$V = \pi \int_a^b R(x)^2 \, dx. \tag{7.2.11}$$

Example 7.2.2: Finding volume using the Disk Method

Find the volume of the solid formed by revolving the curve $y = 1/x$, from $x = 1$ to $x = 2$, around the x-axis.

Solution

A sketch can help us understand this problem. In Figure 7.2.4a the curve $y = 1/x$ is sketched along with the differential element -- a disk -- at x with radius $R(x) = 1/x$. In Figure 7.2.4b the whole solid is pictured, along with the differential element.

The volume of the differential element shown in part (a) of the figure is approximately $\pi R(x_i)^2 \Delta x$, where $R(x_i)$ is the radius of the disk shown and Δx is the thickness of that slice. The radius $R(x_i)$ is the distance from the x-axis to the curve, hence $R(x_i) = 1/x_i$.

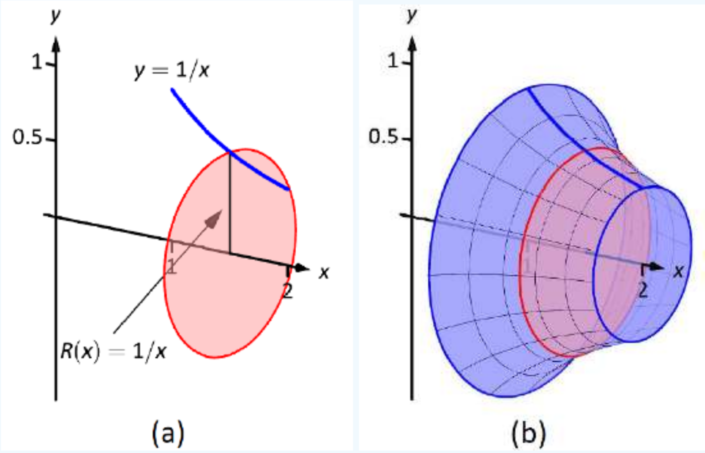


***Figure** 7.2.4: Sketching a solid in Example 7.2.2.*

Slicing the solid into n equally--spaced slices, we can approximate the total volume by adding up the approximate volume of each slice:

$$\text{Approximate volume } = \sum_{i=1}^{n} \pi \left(\frac{1}{x_i}\right)^2 \Delta x. \tag{7.2.12}$$

Taking the limit of the above sum as $n \to \infty$ gives the actual volume; recognizing this sum as a Riemann sum allows us to evaluate the limit with a definite integral, which matches the formula given in Key Idea 23:

$$\begin{aligned} V &= \lim_{n\to\infty} \sum_{i=1}^{n} \pi \left(\frac{1}{x_i}\right)^2 \Delta x && (7.2.13)\\ &= \pi \int_1^2 \left(\frac{1}{x}\right)^2 \, dx && (7.2.14)\\ &= \pi \int_1^2 \frac{1}{x^2} \, dx && (7.2.15)\\ &= \pi \left[-\frac{1}{x}\right]\Big|_1^2 && (7.2.16)\\ &= \pi \left[-\frac{1}{2} - (-1)\right] && (7.2.17)\\ &= \frac{\pi}{2} \text{ units}^3. && (7.2.18) \end{aligned}$$

While Key Idea 23 is given in terms of functions of x, the principle involved can be applied to functions of y when the axis of rotation is vertical, not horizontal. We demonstrate this in the next example.

Example $7.2.3$: Finding volume using the Disk Method

Find the volume of the solid formed by revolving the curve $y = 1/x$, from $x = 1$ to $x = 2$, about the y-axis.

Solution

Since the axis of rotation is vertical, we need to convert the function into a function of y and convert the x-bounds to y-bounds. Since $y = 1/x$ defines the curve, we rewrite it as $x = 1/y$. The bound $x = 1$ corresponds to the y-bound $y = 1$, and the bound $x = 2$ corresponds to the y-bound $y = 1/2$.

Thus we are rotating the curve $x = 1/y$, from $y = 1/2$ to $y = 1$ about the y-axis to form a solid. The curve and sample differential element are sketched in Figure $7.2.5a$, with a full sketch of the solid in Figure $7.2.5b$.

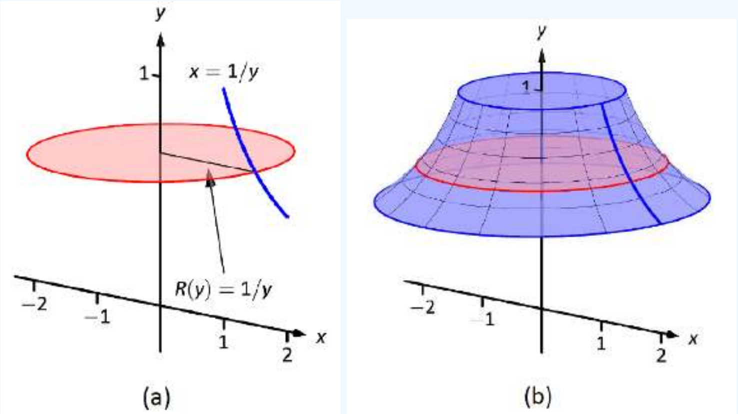


***Figure** $7.2.5$**:** Sketching a solid in Example $7.2.3$.*

We integrate to find the volume:

$$\begin{aligned} V &= \pi \int_{1/2}^{1} \frac{1}{y^2} \, dy && (7.2.19)\\ &= -\frac{\pi}{y}\Big|_{1/2}^{1} && (7.2.20)\\ &= \pi \text{ units}^3. && (7.2.21) \end{aligned}$$

We can also compute the volume of solids of revolution that have a hole in the center. The general principle is simple: compute the volume of the solid irrespective of the hole, then subtract the volume of the hole. If the outside radius of the solid is $R(x)$ and the inside radius (defining the hole) is $r(x)$, then the volume is

$$V = \pi \int^{b} R(x)^2 \, dx - \pi \int^{b} r(x)^2 \, dx = \pi \int^{b} \left(R(x)^2 - r(x)^2\right) \, dx. \qquad (7.2.22)$$

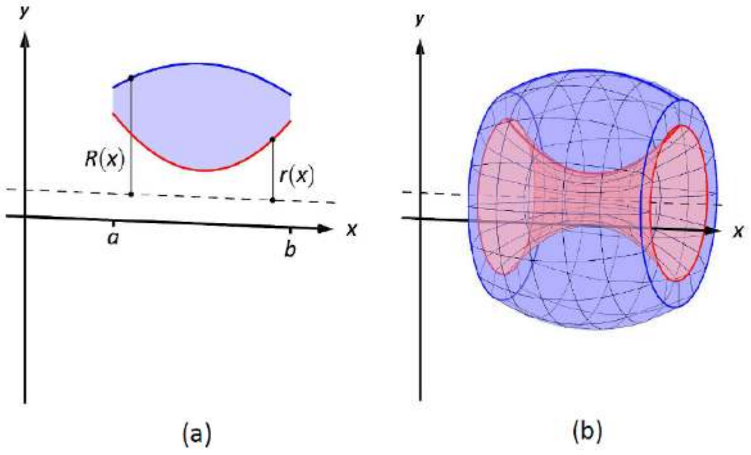


__Figure__ 7.2.6__:__ Establishing the Washer Method; see also Figure 7.2.7.

One can generate a solid of revolution with a hole in the middle by revolving a region about an axis. Consider Figure $7.2.6a$, where a region is sketched along with a dashed, horizontal axis of rotation. By rotating the region about the axis, a solid is formed as sketched in Figure $7.2.6b$. The outside of the solid has radius $R(x)$, whereas the inside has radius $r(x)$. Each cross section of this solid will be a washer (a disk with a hole in the center) as sketched in Figure $7.2.6b$. This leads us to the Washer Method.

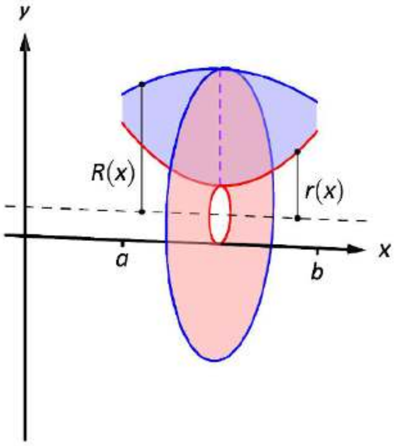


__Figure__ 7.2.7__:__ Establishing the Washer Method; see also Figure 7.2.6.

Key Idea 24: The Washer Method

Let a region bounded by $y = f(x)$, $y = g(x)$, $x = a$ and $x = b$ be rotated about a horizontal axis that does not intersect the region, forming a solid. Each cross section at x will be a washer with outside radius $R(x)$ and inside radius $r(x)$. The volume of the solid is

$$V = \pi \int_a^b \left(R(x)^2 - r(x)^2\right)\, dx. \tag{7.2.23}$$

Even though we introduced it first, the Disk Method is just a special case of the Washer Method with an inside radius of $r(x) = 0$.

Example 7.2.4: Finding volume with the Washer Method

Find the volume of the solid formed by rotating the region bounded by $y = x^2 - 2x + 2$ and $y = 2x - 1$ about the x-axis.

Solution

A sketch of the region will help, as given in Figure $7.2.8a$.

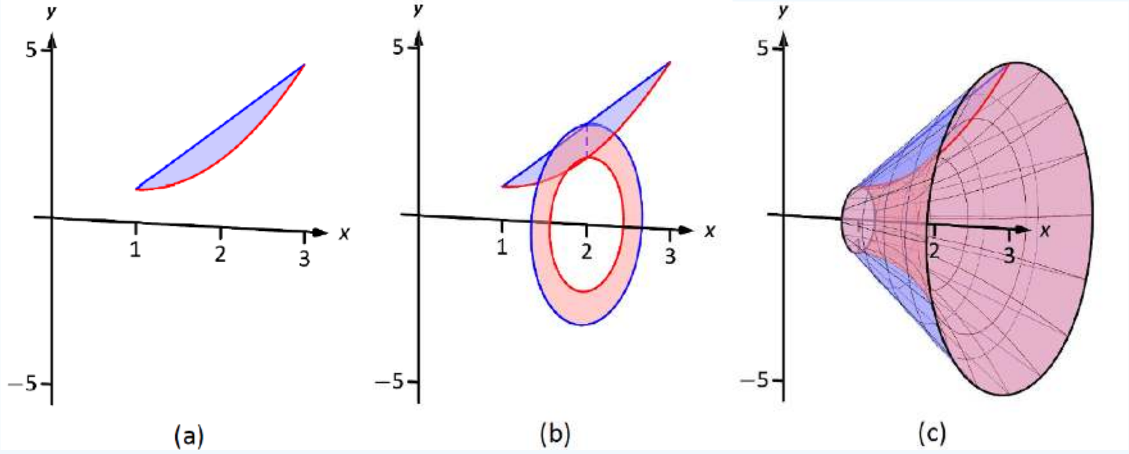


Figure 7.2.8*: Sketching the differential element and solid in Example* 7.2.4.

Rotating about the x-axis will produce cross sections in the shape of washers, as shown in Figure $7.2.8b$; the complete solid is shown in part (c). The outside radius of this washer is $R(x) = 2x+1$; the inside radius is $r(x) = x^2-2x+2$. As the region is bounded from $x=1$ to $x=3$, we integrate as follows to compute the volume.

$$\begin{aligned} V &= \pi\int_1^3 \Big((2x-1)^2-(x^2-2x+2)^2\Big)\,dx && (7.2.24)\\ &= \pi\int_1^3 \Big(-x^4+4x^3-4x^2+4x-3\Big)\,dx && (7.2.25)\\ &= \pi\Big[-\frac{1}{5}x^5+x^4-\frac{4}{3}x^3+2x^2-3x\Big]\Big|_1^3 && (7.2.26)\\ &= \frac{104}{15}\pi \approx 21.78 \text{ units}^3. && (7.2.27)\end{aligned}$$

When rotating about a vertical axis, the outside and inside radius functions must be functions of y.

Example 7.2.5: Finding volume with the Washer Method

Find the volume of the solid formed by rotating the triangular region with vertices at $(1,1)$, $(2,1)$ and $(2,3)$ about the y-axis.

Solution

The triangular region is sketched in Figure 7.2.9; the differential element is sketched in (b) and the full solid is drawn in (c). They help us establish the outside and inside radii. Since the axis of rotation is vertical, each radius is a function of y.

The outside radius $R(y)$ is formed by the line connecting $(2,1)$ and $(2,3)$; it is a constant function, as regardless of the y-value the distance from the line to the axis of rotation is 2. Thus $R(y) = 2$.

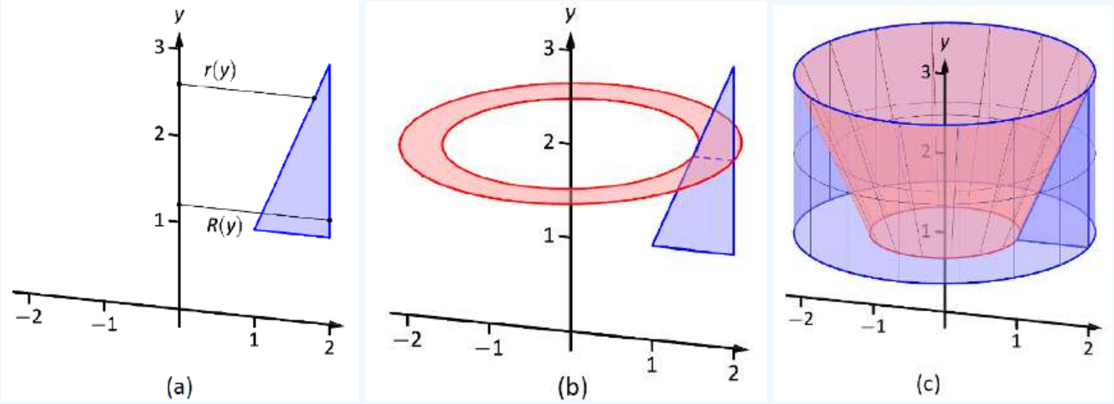


Figure 7.2.9*: Sketching the solid in Example* 7.2.5.

The inside radius is formed by the line connecting $(1, 1)$ and $(2, 3)$. The equation of this line is $y = 2x - 1$, but we need to refer to it as a function of y. Solving for x gives $r(y) = \frac{1}{2}(y + 1)$.

We integrate over the y-bounds of $y = 1$ to $y = 3$. Thus the volume is

$$\begin{aligned} V &= \pi \int_1^3 \left(2^2 - \left(\frac{1}{2}(y+1)\right)^2\right) dy \qquad (7.2.28) \\ &= \pi \int_1^3 \left(-\frac{1}{4}y^2 - \frac{1}{2}y + \frac{15}{4}\right) dy \qquad (7.2.29) \\ &= \pi \left[-\frac{1}{12}y^3 - \frac{1}{4}y^2 + \frac{15}{4}y\right]\Big|_1^3 \qquad (7.2.30) \\ &= \frac{10}{3}\pi \approx 10.47 \text{ units}^3. \qquad (7.2.31) \end{aligned}$$

This section introduced a new application of the definite integral. Our default view of the definite integral is that it gives "the area under the curve." However, we can establish definite integrals that represent other quantities; in this section, we computed volume.

The ultimate goal of this section is not to compute volumes of solids. That can be useful, but what is more useful is the understanding of this basic principle of integral calculus, outlined in Key Idea 24: to find the exact value of some quantity,

- we start with an approximation (in this section, slice the solid and approximate the volume of each slice),
- then make the approximation better by refining our original approximation (i.e., use more slices),
- then use limits to establish a definite integral which gives the exact value.

We practice this principle in the next section where we find volumes by slicing solids in a different way.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/
- Integrated by Justin Marshall.

This page titled 7.2: Volume by Cross-Sectional Area- Disk and Washer Methods is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

7.3: The Shell Method

Often a given problem can be solved in more than one way. A particular method may be chosen out of convenience, personal preference, or perhaps necessity. Ultimately, it is good to have options.

The previous section introduced the Disk and Washer Methods, which computed the volume of solids of revolution by integrating the cross--sectional area of the solid. This section develops another method of computing volume, the **Shell Method**. Instead of slicing the solid perpendicular to the axis of rotation creating cross-sections, we now slice it parallel to the axis of rotation, creating "shells."

Consider Figure $7.3.1$, where the region shown in (a) is rotated around the y-axis forming the solid shown in (b). A small slice of the region is drawn in (a), parallel to the axis of rotation. When the region is rotated, this thin slice forms a **cylindrical shell**, as pictured in part (c) of the figure. The previous section approximated a solid with lots of thin disks (or washers); we now approximate a solid with many thin cylindrical shells.

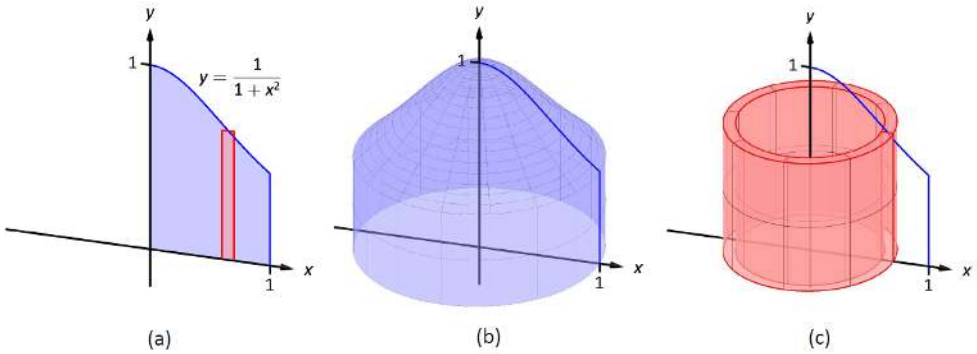


Figure $7.3.1$**:** Introducing the Shell Method.

Figure $7.3.1$ **(d):** A dynamic version of this figure created using CalcPlot3D.

To compute the volume of one shell, first consider the paper label on a soup can with radius r and height h. What is the area of this label? A simple way of determining this is to cut the label and lay it out flat, forming a rectangle with height h and length $2\pi r$. Thus the area is $A = 2\pi rh$; see Figure $7.3.2a$.

Do a similar process with a cylindrical shell, with height h, thickness Δx, and approximate radius r. Cutting the shell and laying it flat forms a rectangular solid with length $2\pi r$, height h and depth dx. Thus the volume is $V \approx 2\pi rh\ dx$; see Figure $7.3.2c$. (We say "approximately" since our radius was an approximation.)

By breaking the solid into n cylindrical shells, we can approximate the volume of the solid as

$$V = \sum_{i=1}^{n} 2\pi r_i h_i\ dx_i, \tag{7.3.1}$$

where r_i, h_i and dx_i are the radius, height and thickness of the $i\,^{\text{th}}$ shell, respectively.

This is a Riemann Sum. Taking a limit as the thickness of the shells approaches 0 leads to a definite integral.

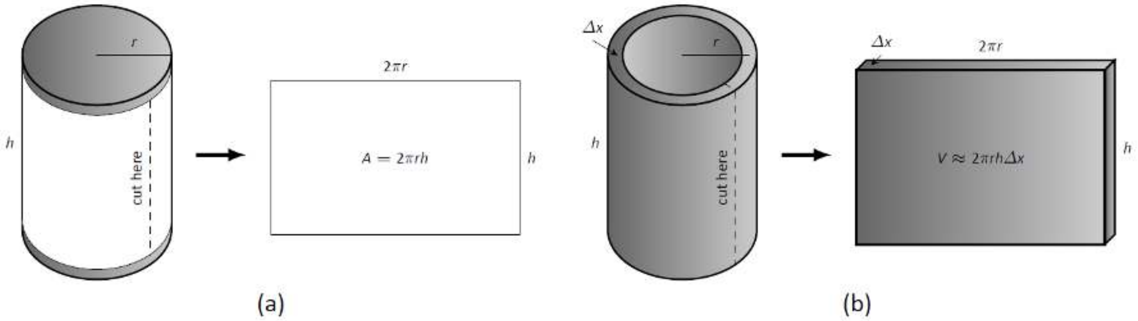


Figure $7.3.2$***:*** *Determining the volume of a thin cylindrical shell.}\label{fig:soupcan}*

Key Idea 25: Shell Method

Let a solid be formed by revolving a region R, bounded by $x = a$ and $x = b$, around a vertical axis. Let $r(x)$ represent the distance from the axis of rotation to x (i.e., the radius of a sample shell) and let $h(x)$ represent the height of the solid at x (i.e., the height of the shell). The volume of the solid is

$$V = 2\pi \int_a^b r(x)h(x)\ dx. \tag{7.3.2}$$

Special Cases:

1. When the region R is bounded above by $y = f(x)$ and below by $y = g(x)$, then $h(x) = f(x) - g(x)$.
2. When the axis of rotation is the y-axis (i.e., $x = 0$) then $r(x) = x$.

Let's practice using the Shell Method.

Example 7.3.1: Finding volume using the Shell Method

Find the volume of the solid formed by rotating the region bounded by $y = 0$, $y = 1/(1+x^2)$, $x = 0$ and $x = 1$ about the y-axis.

Solution

This is the region used to introduce the Shell Method in Figure 7.3.1, but is sketched again in Figure 7.3.3 for closer reference. A line is drawn in the region parallel to the axis of rotation representing a shell that will be carved out as the region is rotated about the y-axis. (This is the differential element.)

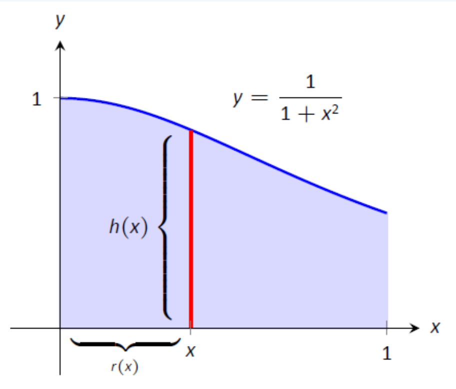


***Figure** 7.3.3**:** Graphing a region in Example 7.3.1.*

The distance this line is from the axis of rotation determines $r(x)$; as the distance from x to the y-axis is x, we have $r(x) = x$. The height of this line determines $h(x)$; the top of the line is at $y = 1/(1+x^2)$, whereas the bottom of the line is at $y = 0$. Thus $h(x) = 1/(1+x^2) - 0 = 1/(1+x^2)$. The region is bounded from $x = 0$ to $x = 1$, so the volume is

$$V = 2\pi \int_0^1 \frac{x}{1+x^2}\ dx. \tag{7.3.3}$$

This requires substitution. Let $u = 1 + x^2$, so $du = 2x\ dx$. We also change the bounds: $u(0) = 1$ and $u(1) = 2$. Thus we have:

$$\begin{aligned} &= \pi \int_1^2 \frac{1}{u}\ du \\ &= \pi \ln u \Big|_1^2 \\ &= \pi \ln 2 - \pi \ln 1 \\ &= \pi \ln 2 \approx 2.178 \text{ units}^3. \end{aligned}$$

Note: in order to find this volume using the Disk Method, two integrals would be needed to account for the regions above and below $y = 1/2$.

With the Shell Method, nothing special needs to be accounted for to compute the volume of a solid that has a hole in the middle, as demonstrated next.

Example $7.3.2$: Finding volume using the Shell Method

Find the volume of the solid formed by rotating the triangular region determined by the points $(0,1)$, $(1,1)$ and $(1,3)$ about the line $x=3$.

Solution

The region is sketched in Figure $7.3.4a$ along with the differential element, a line within the region parallel to the axis of rotation. In part (b) of the figure, we see the shell traced out by the differential element, and in part (c) the whole solid is shown.

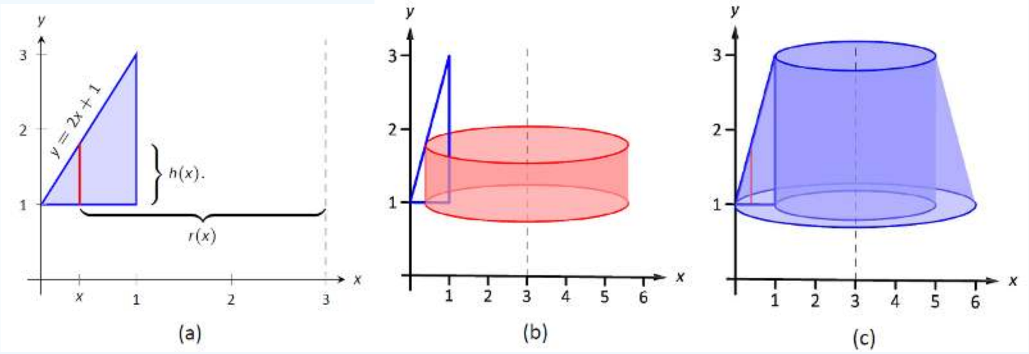


***Figure** $7.3.4$**:** Graphing a region in Example $7.3.2$*

The height of the differential element is the distance from $y=1$ to $y=2x+1$, the line that connects the points $(0,1)$ and $(1,3)$. Thus $h(x)=2x+1-1=2x$. The radius of the shell formed by the differential element is the distance from x to $x=3$; that is, it is $r(x)=3-x$. The x-bounds of the region are $x=0$ to $x=1$, giving

$$\begin{aligned} V &= 2\pi\int_0^1 (3-x)(2x)\ dx \\ &= 2\pi\int_0^1 \left(6x-2x^2\right)\ dx \\ &= 2\pi\left(3x^2-\frac{2}{3}x^3\right)\Big|_0^1 \\ &= \frac{14}{3}\pi \approx 14.66\ \text{units}^3. \end{aligned}$$

When revolving a region around a horizontal axis, we must consider the radius and height functions in terms of y, not x.

Example $7.3.3$: Finding volume using the Shell Method

Find the volume of the solid formed by rotating the region given in Example $7.3.2$ about the x-axis.

Solution

The region is sketched in Figure $7.3.5a$ with a sample differential element. In part (b) of the figure the shell formed by the differential element is drawn, and the solid is sketched in (c). (Note that the triangular region looks "short and wide" here, whereas in the previous example the same region looked "tall and narrow." This is because the bounds on the graphs are different.)

The height of the differential element is an x-distance, between $x=\frac{1}{2}y-\frac{1}{2}$ and $x=1$. Thus $h(y)=1-(\frac{1}{2}y-\frac{1}{2})=-\frac{1}{2}y+\frac{3}{2}$. The radius is the distance from y to the x-axis, so $r(y)=y$. The y bounds of the region are $y=1$ and $y=3$, leading to the integral

$$\begin{aligned} V &= 2\pi \int_1^3 \left[y\left(-\frac{1}{2}y+\frac{3}{2}\right)\right]\, dy \\ &= 2\pi \int_1^3 \left[-\frac{1}{2}y^2+\frac{3}{2}y\right]\, dy \\ &= 2\pi\left[-\frac{1}{6}y^3+\frac{3}{4}y^2\right]\Big|_1^3 \\ &= 2\pi\left[\frac{9}{4}-\frac{7}{12}\right] \\ &= \frac{10}{3}\pi \approx 10.472 \text{ units}^3. \end{aligned}$$

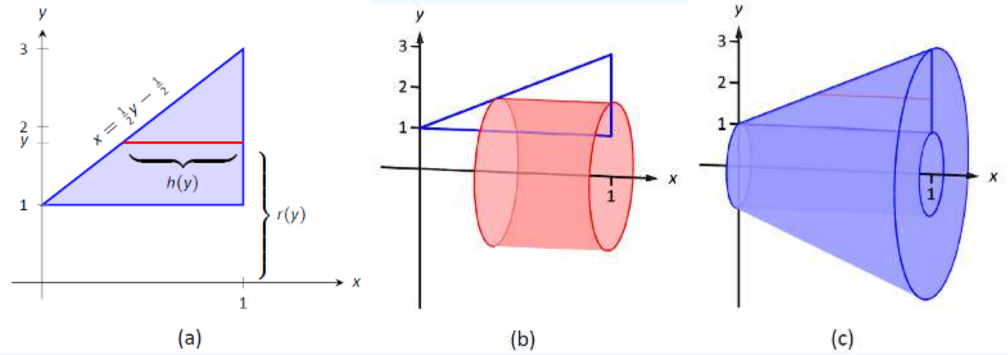


***Figure** 7.3.5**:** Graphing a region in Example 7.3.3*

At the beginning of this section it was stated that "it is good to have options." The next example finds the volume of a solid rather easily with the Shell Method, but using the Washer Method would be quite a chore.

Example 7.3.4: Finding volume using the Shell Method

Find the volume of the solid formed by revolving the region bounded by $y = \sin x$ and the x-axis from $x = 0$ to $x = \pi$ about the y-axis.

Solution

The region and a differential element, the shell formed by this differential element, and the resulting solid are given in Figure 7.3.6.

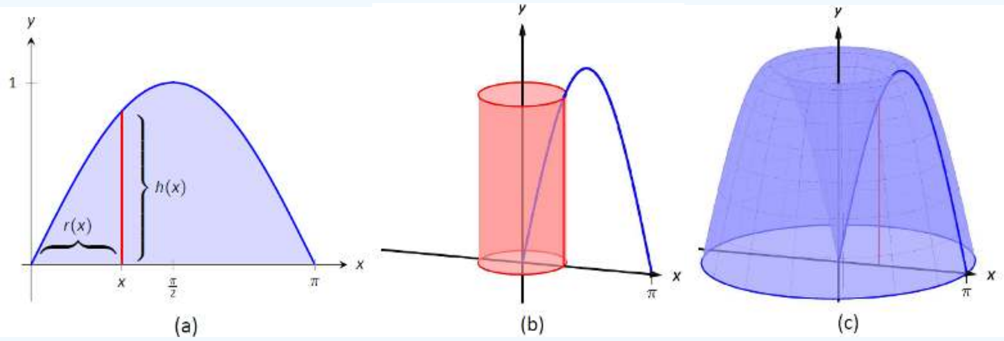


***Figure** 7.3.6**:** Graphing a region in Example 7.3.4*

The radius of a sample shell is $r(x) = x$; the height of a sample shell is $h(x) = \sin x$, each from $x = 0$ to $x = \pi$. Thus the volume of the solid is

$$V = 2\pi \int_0^\pi x \sin x \, dx. \tag{7.3.4}$$

This requires Integration By Parts. Set $u = x$ and $dv = \sin x\, dx$; we leave it to the reader to fill in the rest. We have:

$$\begin{aligned} &= 2\pi\Big[-x\cos x\Big|_0^{\pi} + \int_0^{\pi} \cos x \; dx\Big] \\ &= 2\pi\Big[\pi + \sin x\Big|_0^{\pi}\Big] \\ &= 2\pi\Big[\pi + 0\Big] \\ &= 2\pi^2 \approx 19.74 \text{ units}^3. \end{aligned}$$

Note that in order to use the Washer Method, we would need to solve $y = \sin x$ for x, requiring the use of the arcsine function. We leave it to the reader to verify that the outside radius function is $R(y) = \pi - \arcsin y$ and the inside radius function is $r(y) = \arcsin y$. Thus the volume can be computed as

$$\pi \int_0^1 \Big[(\pi - \arcsin y)^2 - (\arcsin y)^2\Big] \; dy. \tag{7.3.5}$$

This integral isn't terrible given that the $\arcsin^2 y$ terms cancel, but it is more onerous than the integral created by the Shell Method.

We end this section with a table summarizing the usage of the Washer and Shell Methods.

Key Idea 26: Summary of the Washer and Shell Methods

Let a region R be given with x-bounds $x = a$ and $x = b$ and y-bounds $y = c$ and $y = d$.

	Washer Method	Shell Method
Horizontal Axis	$\pi \int_a^b \left(R(x)^2 - r(x)^2\right) \; dx$	$2\pi \int_c^d r(y)h(y) \; dy$
Vertical Axis	$\pi \int_c^d \left(R(y)^2 - r(y)^2\right) \; dy$	$2\pi \int_a^b r(x)h(x) \; dx$

As in the previous section, the real goal of this section is not to be able to compute volumes of certain solids. Rather, it is to be able to solve a problem by first approximating, then using limits to refine the approximation to give the exact value. In this section, we approximate the volume of a solid by cutting it into thin cylindrical shells. By summing up the volumes of each shell, we get an approximation of the volume. By taking a limit as the number of equally spaced shells goes to infinity, our summation can be evaluated as a definite integral, giving the exact value.

We use this same principle again in the next section, where we find the length of curves in the plane.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/
- Integrated by Justin Marshall.

This page titled 7.3: The Shell Method is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

7.4: Arc Length and Surface Area

Arc Length

In previous sections we have used integration to answer the following questions:

1. Given a region, what is its area?
2. Given a solid, what is its volume?

In this section, we address a related question: Given a curve, what is its length? This is often referred to as **arc length**.

Consider the graph of $y=\sin x$ on $[0,\pi]$ given in Figure $7.4.1a$. How long is this curve? That is, if we were to use a piece of string to exactly match the shape of this curve, how long would the string be?

As we have done in the past, we start by approximating; later, we will refine our answer using limits to get an exact solution.

The length of straight--line segments is easy to compute using the Distance Formula. We can approximate the length of the given curve by approximating the curve with straight lines and measuring their lengths.

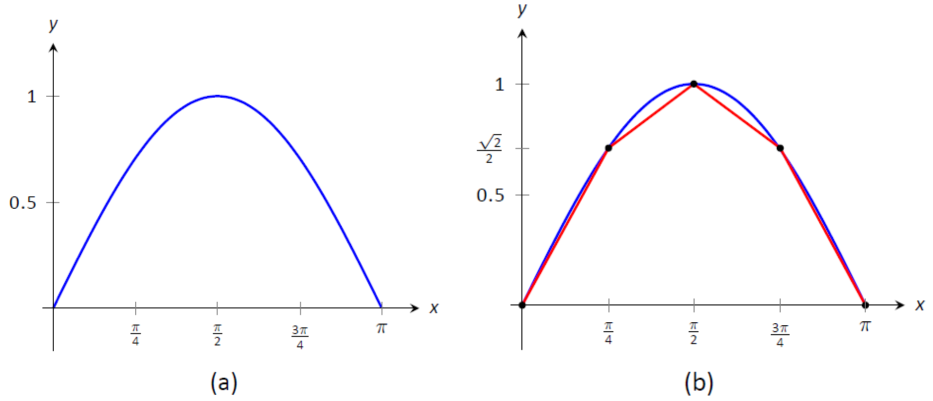


***Figure** $7.4.1$**:** Graphing $y=\sin x$ on $[0,\pi]$ and approximating the curve with line segments.*

In Figure $7.4.1b$, the curve $y=\sin x$ has been approximated with 4 line segments (the interval $[0,\pi]$ has been divided into 4 equally--lengthed subintervals). It is clear that these four line segments approximate $y=\sin x$ very well on the first and last subinterval, though not so well in the middle. Regardless, the sum of the lengths of the line segments is 3.79, so we approximate the arc length of $y=\sin x$ on $[0,\pi]$ to be 3.79.

In general, we can approximate the arc length of $y=f(x)$ on $[a,b]$ in the following manner. Let $a=x_1<x_2<\ldots<x_n<x_{n+1}=b$ be a partition of $[a,b]$ into n subintervals. Let dx_i represent the length of the i^{th} subinterval $[x_i,x_{i+1}]$.

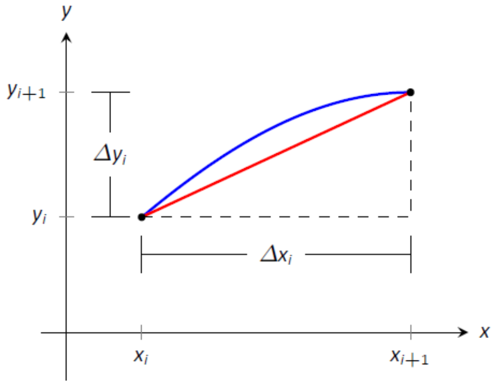


***Figure** $7.4.2$**:** Zooming in on the i^{th} subinterval $[x_i,x_{i+1}]$ of a partition of $[a,b]$.*

Figure $7.4.2$ zooms in on the i^{th} subinterval where $y=f(x)$ is approximated by a straight line segment. The dashed lines show that we can view this line segment as they hypotenuse of a right triangle whose sides have length dx_i and dy_i. Using the Pythagorean

Theorem, the length of this line segment is $\sqrt{dx_i^2+\Delta y_i^2}$. Summing over all subintervals gives an arc length approximation

$$L \approx \sum_{i=1}^{n} \sqrt{dx_i^2+\Delta y_i^2}. \tag{7.4.1}$$

As shown here, this is *not* a Riemann Sum. While we could conclude that taking a limit as the subinterval length goes to zero gives the exact arc length, we would not be able to compute the answer with a definite integral. We need first to do a little algebra.

In the above expression factor out a dx_i^2 term:

$$\sum_{i=1}^{n} \sqrt{dx_i^2+\Delta y_i^2} = \sum_{i=1}^{n} \sqrt{dx_i^2\left(1+\frac{\Delta y_i^2}{dx_i^2}\right)}. \tag{7.4.2}$$

Now pull the dx_i^2 term out of the square root:

$$= \sum_{i=1}^{n} \sqrt{1+\frac{\Delta y_i^2}{dx_i^2}}\ dx_i. \tag{7.4.3}$$

This is nearly a Riemann Sum. Consider the $\Delta y_i^2/dx_i^2$ term. The expression $\Delta y_i/dx_i$ measures the "change in y/change in x," that is, the "rise over run" of f on the i^{th} subinterval. The Mean Value Theorem of Differentiation (Theorem 3.2.1) states that there is a c_i in the i^{th} subinterval where $f'(c_i)=\Delta y_i/dx_i$. Thus we can rewrite our above expression as:

$$= \sum_{i=1}^{n} \sqrt{1+f'(c_i)^2}\ dx_i. \tag{7.4.4}$$

This *is* a Riemann Sum. As long as f' is continuous, we can invoke Theorem 5.3.2 and conclude

$$= \int_a^b \sqrt{1+f'(x)^2}\ dx. \tag{7.4.5}$$

Key Idea 27: Arc Length

Let f be differentiable on an open interval containing $[a,b]$, where f' is also continuous on $[a,b]$. Then the arc length of f from $x=a$ to $x=b$ is

$$L = \int_a^b \sqrt{1+f'(x)^2}\ dx. \tag{7.4.6}$$

As the integrand contains a square root, it is often difficult to use the formula in Key Idea 27 to find the length exactly. When exact answers are difficult to come by, we resort to using numerical methods of approximating definite integrals. The following examples will demonstrate this.

Example 7.4.1: Finding arc length

Find the arc length of $f(x)=x^{3/2}$ from $x=0$ to $x=4$.

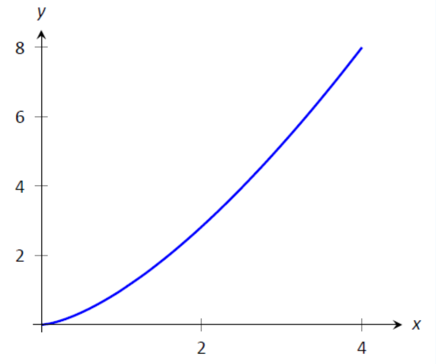


***Figure** 7.4.3: A graph of $f(x)=x^{3/2}$ from Example 7.4.1*

Solution

We begin by finding $f'(x) = \frac{3}{2}x^{1/2}$. Using the formula, we find the arc length L as

$$
\begin{aligned}
L &= \int_0^4 \sqrt{1+\left(\frac{3}{2}x^{1/2}\right)^2}\,dx && (7.4.7)\\
&= \int_0^4 \sqrt{1+\frac{9}{4}x}\,dx && (7.4.8)\\
&= \int_0^4 \left(1+\frac{9}{4}x\right)^{1/2}\,dx && (7.4.9)\\
&= \frac{2}{3}\frac{4}{9}\left(1+\frac{9}{4}x\right)^{3/2}\Big|_0^4 && (7.4.10)\\
&= \frac{8}{27}\left(10^{3/2}-1\right) \approx 9.07\text{units.} && (7.4.11)
\end{aligned}
$$

Example 7.4.2: Finding arc length

Find the arc length of $f(x) = \frac{1}{8}x^2 - \ln x$ from $x = 1$ to $x = 2$.

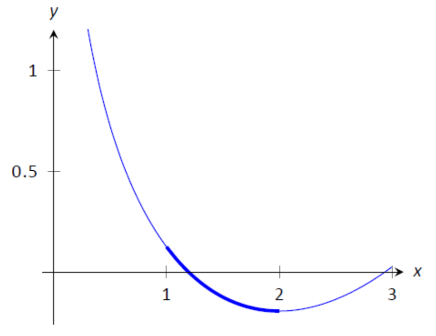


***Figure** 7.4.4: A graph of $f(x) = \frac{1}{8}x^2 - \ln x$ from Example 7.4.2.*

Solution

This function was chosen specifically because the resulting integral can be evaluated exactly. We begin by finding $f'(x) = x/4 - 1/x$. The arc length is

$$
\begin{aligned}
L &= \int_1^2 \sqrt{1+\left(\frac{x}{4}-\frac{1}{x}\right)^2}\,dx && (7.4.12)\\
&= \int_1^2 \sqrt{1+\frac{x^2}{16}-\frac{1}{2}+\frac{1}{x^2}}\,dx && (7.4.13)\\
&= \int_1^2 \sqrt{\frac{x^2}{16}+\frac{1}{2}+\frac{1}{x^2}}\,dx && (7.4.14)\\
&= \int_1^2 \sqrt{\left(\frac{x}{4}+\frac{1}{x}\right)^2}\,dx \qquad = \int_1^2 \left(\frac{x}{4}+\frac{1}{x}\right)\,dx && (7.4.15)\\
&= \left(\frac{x^2}{8}+\ln x\right)\Big|_1^2 && (7.4.16)\\
&= \frac{3}{8}+\ln 2 \approx 1.07\text{ units.} && (7.4.17)
\end{aligned}
$$

A graph of f is given in Figure 7.4.4; the portion of the curve measured in this problem is in bold.

The previous examples found the arc length exactly through careful choice of the functions. In general, exact answers are much more difficult to come by and numerical approximations are necessary.

Example 7.4.3: Approximating arc length numerically

Find the length of the sine curve from $x=0$ to $x=\pi$.

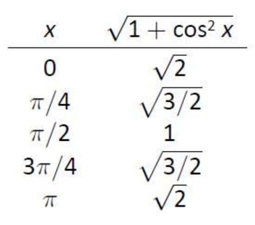
Solution

This is somewhat of a mathematical curiosity; in Example 5.4.3 we found the area under one "hump" of the sine curve is 2 square units; now we are measuring its arc length.

The setup is straightforward: $f(x)=\sin x$ and $f'(x)=\cos x$. Thus

$$L=\int_0^\pi \sqrt{1+\cos^2 x}\;dx. \tag{7.4.18}$$

This integral *cannot* be evaluated in terms of elementary functions so we will approximate it with Simpson's Method with $n=4$.



x	$\sqrt{1+\cos^2 x}$
0	$\sqrt{2}$
$\pi/4$	$\sqrt{3/2}$
$\pi/2$	1
$3\pi/4$	$\sqrt{3/2}$
π	$\sqrt{2}$

***Figure** 7.4.5: A table of values of $y=\sqrt{1+\cos^2 x}$ to evaluate a definite integral in Example 7.4.3.*

$$\begin{array}{cc} x & \sqrt{1+\cos^2 x} \\ \hline 0 & \sqrt{2} \\ \pi/4 & \sqrt{3/2} \\ \pi/2 & 1 \\ 3\pi/4 & \sqrt{3/2} \\ \pi & \sqrt{2} \end{array} \tag{7.4.19}$$

Figure ??? gives $\sqrt{1+\cos^2 x}$ evaluated at 5 evenly spaced points in $[0,\pi]$. Simpson's Rule then states that

$$\int_0^\pi \sqrt{1+\cos^2 x}\;dx \approx \frac{\pi-0}{4\cdot 3}\left(\sqrt{2}+4\sqrt{3/2}+2(1)+4\sqrt{3/2}+\sqrt{2}\right) \tag{7.4.20}$$

$$=3.82918. \tag{7.4.21}$$

Using a computer with $n=100$ the approximation is $L\approx 3.8202$; our approximation with $n=4$ is quite good.

Surface Area of Solids of Revolution

We have already seen how a curve $y=f(x)$ on $[a,b]$ can be revolved around an axis to form a solid. Instead of computing its volume, we now consider its surface area.

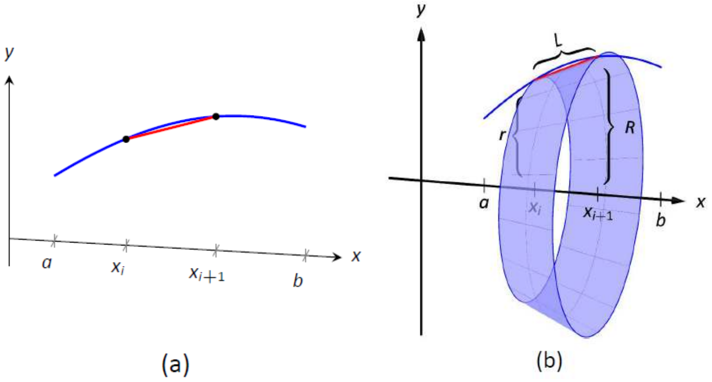


***Figure** 7.4.6: Establishing the formula for surface area.*

We begin as we have in the previous sections: we partition the interval $[a,b]$ with n subintervals, where the $i^{\text{ th}}$ subinterval is $[x_i, x_{i+1}]$. On each subinterval, we can approximate the curve $y=f(x)$ with a straight line that connects $f(x_i)$ and $f(x_{i+1})$ as shown in Figure $7.4.5a$. Revolving this line segment about the x-axis creates part of a cone (called a *frustum* of a cone) as shown in Figure $7.4.5b$. The surface area of a frustum of a cone is

$$2\pi\cdot \text{ length } \cdot \text{average of the two radii } R \text{ and } r. \tag{7.4.22}$$

The length is given by L; we use the material just covered by arc length to state that

$$L \approx \sqrt{1+f'(c_i)}dx_i \tag{7.4.23}$$

for some c_i in the $i^{\text{ th}}$ subinterval. The radii are just the function evaluated at the endpoints of the interval. That is,

$$R=f(x_{i+1}) \quad \text{and} \quad r=f(x_i). \tag{7.4.24}$$

Thus the surface area of this sample frustum of the cone is approximately

$$2\pi\frac{f(x_i)+f(x_{i+1})}{2}\sqrt{1+f'(c_i)^2}dx_i. \tag{7.4.25}$$

Since f is a continuous function, the Intermediate Value Theorem states there is some d_i in $[x_i, x_{i+1}]$ such that $f(d_i)=\frac{f(x_i)+f(x_{i+1})}{2}$; we can use this to rewrite the above equation as

$$2\pi f(d_i)\sqrt{1+f'(c_i)^2}dx_i. \tag{7.4.26}$$

Summing over all the subintervals we get the total surface area to be approximately

$$\text{Surface Area} \approx \sum_{i=1}^{n} 2\pi f(d_i)\sqrt{1+f'(c_i)^2}dx_i, \tag{7.4.27}$$

which is a Riemann Sum. Taking the limit as the subinterval lengths go to zero gives us the exact surface area, given in the following Key Idea.

Key Idea 28: Surface Area of a Solid of Revolution

Let f be differentiable on an open interval containing $[a,b]$ where f' is also continuous on $[a,b]$.

1. The surface area of the solid formed by revolving the graph of $y=f(x)$, where $f(x)\geq 0$, about the x-axis is

$$\text{Surface Area} = 2\pi\int_a^b f(x)\sqrt{1+f'(x)^2}\ dx. \tag{7.4.28}$$

2. The surface area of the solid formed by revolving the graph of $y=f(x)$ about the y-axis, where $a, b\geq 0$, is

$$\text{Surface Area} = 2\pi\int_a^b x\sqrt{1+f'(x)^2}\ dx. \tag{7.4.29}$$

When revolving $y=f(x)$ about the y-axis, the radii of the resulting frustum are x_i and x_{i+1}; their average value is simply the midpoint of the interval. In the limit, this midpoint is just x. This gives the second part of Key Idea 28.

Example $7.4.4$: Finding surface area of a solid of revolution

Find the surface area of the solid formed by revolving $y=\sin x$ on $[0,\pi]$ around the x-axis, as shown in Figure $7.4.6$.

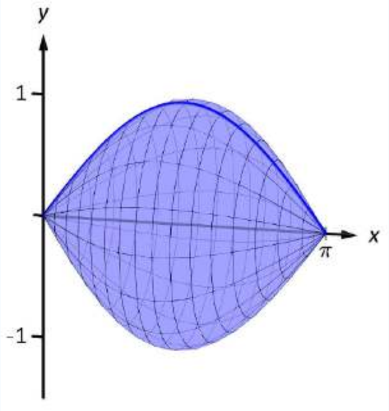


Figure $7.4.7$***:*** *Revolving* $y = \sin x$ *on* $[0, \pi]$ *about the* x*-axis.*

Solution

The setup is relatively straightforward. Using Key Idea ???, we have the surface area SA is:

$$SA = 2\pi \int_0^{\pi} \sin x \sqrt{1+\cos^2 x}\; dx \tag{7.4.30}$$

$$= -2\pi \frac{1}{2}\left(\sinh^{-1}(\cos x) + \cos x\sqrt{1+\cos^2 x}\right)\Big|_0^{\pi} \tag{7.4.31}$$

$$= 2\pi\left(\sqrt{2} + \sinh^{-1} 1\right) \tag{7.4.32}$$

$$\approx 14.42 \text{ units}^2. \tag{7.4.33}$$

The integration step above is nontrivial, utilizing an integration method called Trigonometric Substitution.

It is interesting to see that the surface area of a solid, whose shape is defined by a trigonometric function, involves both a square root and an inverse hyperbolic trigonometric function.

Example $7.4.5$: Finding surface area of a solid of revolution

Find the surface area of the solid formed by revolving the curve $y = x^2$ on $[0, 1]$ about:

1. the x-axis
2. the y-axis.

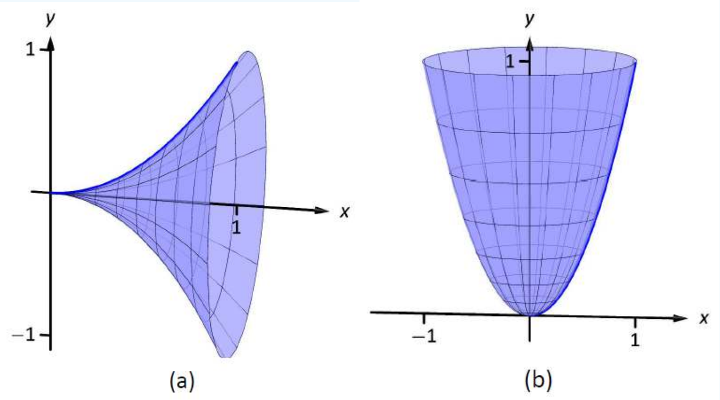


Figure $7.4.8$***:*** *The solids used in Example* $7.4.6$*.*

Solution

1. The integral is straightforward to setup:

$$SA = 2\pi \int_0^1 x^2 \sqrt{1+(2x)^2} dx. \tag{7.4.34}$$

Like the integral in Example ???, this requires Trigonometric Substitution.

$$= \frac{\pi}{32}\left(2(8x^3+x)\sqrt{1+4x^2}-\sinh^{-1}(2x)\right)\Big|_0^1 \tag{7.4.35}$$

$$= \frac{\pi}{32}\left(18\sqrt{5}-\sinh^{-1}2\right) \tag{7.4.36}$$

$$\approx 3.81 \text{ units}^2. \tag{7.4.37}$$

The solid formed by revolving $y=x^2$ around the x-axis is graphed in Figure $7.4.7a$.

2. Since we are revolving around the y-axis, the "radius" of the solid is not $f(x)$ but rather x. Thus the integral to compute the surface area is:

$$SA = 2\pi\int_0^1 x\sqrt{1+(2x)^2}dx. \tag{7.4.38}$$

This integral can be solved using substitution. Set $u=1+4x^2$; the new bounds are $u=1$ to $u=5$. We then have

$$= \frac{\pi}{4}\int_1^5 \sqrt{u}du \tag{7.4.39}$$

$$= \frac{\pi}{4}\frac{2}{3}u^{3/2}\Big|_1^5 \tag{7.4.40}$$

$$= \frac{\pi}{6}\left(5\sqrt{5}-1\right) \tag{7.4.41}$$

$$\approx 5.33 \text{ units}^2. \tag{7.4.42}$$

The solid formed by revolving $y=x^2$ about the y-axis is graphed in Figure $7.4.7b$.

Our final example is a famous mathematical "paradox."

Example $7.4.6$: The surface area and volume of Gabriel's Horn

Consider the solid formed by revolving $y=1/x$ about the x-axis on $[1,\infty)$. Find the volume and surface area of this solid. (This shape, as graphed in Figure $7.4.9$, is known as "Gabriel's Horn" since it looks like a very long horn that only a supernatural person, such as an angel, could play.)

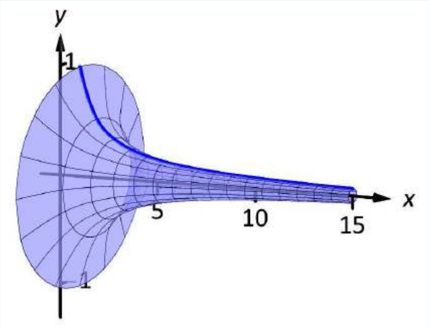


Figure $7.4.9$***:*** *A graph of Gabriel's Horn.*

Solution

To compute the volume it is natural to use the Disk Method. We have:

$$V = \pi\int_1^\infty \frac{1}{x^2}\,dx \tag{7.4.43}$$

$$= \lim_{b\to\infty}\pi\int_1^b \frac{1}{x^2}\,dx \tag{7.4.44}$$

$$= \lim_{b\to\infty}\pi\left(\frac{-1}{x}\right)\Big|_1^b \tag{7.4.45}$$

$$= \lim_{b\to\infty}\pi\left(1-\frac{1}{b}\right) \tag{7.4.46}$$

$$= \pi \text{ units}^3. \tag{7.4.47}$$

Gabriel's Horn has a finite volume of π cubic units. Since we have already seen that regions with infinite length can have a finite area, this is not too difficult to accept.

We now consider its surface area. The integral is straightforward to setup:

$$SA = 2\pi \int_1^\infty \frac{1}{x}\sqrt{1+1/x^4}\;dx. \tag{7.4.48}$$

Integrating this expression is not trivial. We can, however, compare it to other improper integrals. Since $1 < \sqrt{1+1/x^4}$ on $[1,\infty)$, we can state that

$$2\pi \int_1^\infty \frac{1}{x}dx < 2\pi \int_1^\infty \frac{1}{x}\sqrt{1+1/x^4}dx. \tag{7.4.49}$$

By Key Idea 21, the improper integral on the left diverges. Since the integral on the right is larger, we conclude it also diverges, meaning Gabriel's Horn has infinite surface area.

Hence the "paradox": we can fill Gabriel's Horn with a finite amount of paint, but since it has infinite surface area, we can never paint it.

Somehow this paradox is striking when we think about it in terms of volume and area. However, we have seen a similar paradox before, as referenced above. We know that the area under the curve $y = 1/x^2$ on $[1,\infty)$ is finite, yet the shape has an infinite perimeter. Strange things can occur when we deal with the infinite.

A standard equation from physics is "Work = force $\times$ distance", when the force applied is constant. In the next section we learn how to compute work when the force applied is variable.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/
- Integrated by Justin Marshall.

This page titled 7.4: Arc Length and Surface Area is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

7.5: Work

Work is the scientific term used to describe the action of a force which moves an object. When a constant force $\vec{F}$ is applied to move an object a distance d, the amount of work performed is

$$W = \vec{F} \cdot \vec{d}. \tag{7.5.1}$$

The SI unit of force is the Newton, (kg·m/s^2) and the SI unit of distance is a meter (m). The fundamental unit of work is one Newton-meter, or a joule (J). That is, applying a force of one Newton for one meter performs one joule of work. In Imperial units (as used in the United States), force is measured in pounds (lb) and distance is measured in feet (ft), hence work is measured in ft-lb.

Mass vs. Weight

Mass and *weight* are closely related, yet different, concepts. The mass m of an object is a quantitative measure of that object's resistance to acceleration. The weight w of an object is a measurement of the force applied to the object by the acceleration of gravity g.

Since the two measurements are proportional, $w = m \cdot g$, they are often used interchangeably in everyday conversation. When computing work, one must be careful to note which is being referred to. When mass is given, it must be multiplied by the acceleration of gravity to reference the related force.

When force is constant, the measurement of work is straightforward. For instance, lifting a 200 lb object 5 ft performs $200 \cdot 5 = 1000$ ft-lb of work.

What if the force applied is variable? For instance, imagine a climber pulling a 200 ft rope up a vertical face. The rope becomes lighter as more is pulled in, requiring less force and hence the climber performs less work.

In general, let $F(x)$ be a force function on an interval $[a, b]$. We want to measure the amount of work done applying the force F from $x = a$ to $x = b$. We can approximate the amount of work being done by partitioning $[a, b]$ into subintervals $a = x_1 < x_2 < \cdots < x_{n+1} = b$ and assuming that F is constant on each subinterval. Let c_i be a value in the i^{th} subinterval $[x_i, x_{i+1}]$. Then the work done on this interval is approximately

$$W_i \approx F(c_i) \cdot (x_{i+1} - x_i) = F(c_i)\,\Delta x_i,$$

a constant force × the distance over which it is applied. The total work is

$$W = \sum_{i=1}^{n} W_i \approx \sum_{i=1}^{n} F(c_i)\,\Delta x_i.$$

This, of course, is a Riemann sum. Taking a limit as the subinterval lengths go to zero give an exact value of work which can be evaluated through a definite integral.

Key Idea 29: Work

Let $F(x)$ be a continuous function on $[a, b]$ describing the amount of force being applied to an object in the direction of travel from distance $x = a$ to distance $x = b$. The total work W done on $[a, b]$ is

$$W = \int_a^b F(x)\,dx. \tag{7.5.2}$$

Example 7.5.1: Computing work performed - applying variable force

A 60m climbing rope is hanging over the side of a tall cliff. How much work is performed in pulling the rope up to the top, where the rope has a mass of 66g/m? How much work is performed pulling a 60 m climbing rope up a cliff face, where the rope has a mass of 66 g/m?

Solution

We need to create a force function $F(x)$ on the interval $[0, 60]$. To do so, we must first decide what x is measuring: is it the length of the rope still hanging, or is it the amount of rope pulled in? As long as we are consistent, either approach is fine. We adopt for this example the convention that x is the amount of rope pulled in. This seems to match intuition better; pulling up the first 10 meters of rope involves $x = 0$ to $x = 10$ instead of $x = 60$ to $x = 50$.

As x is the amount of rope pulled in, the amount of rope still hanging is $60 - x$. This length of rope has a mass of 66 g/m, or 0.066 kg/m. The the mass of the rope still hanging is $0.066(60 - x)$ kg; multiplying this mass by the acceleration of gravity, 9.8 m/s^2, gives our variable force function

$$F(x) = (9.8)(0.066)(60 - x) = 0.6468(60 - x).$$

Thus the total work performed in pulling up the rope is

$$W = \int_0^{60} 0.6468(60 - x)\,dx = 1,164.24 \text{ J}.$$

By comparison, consider the work done in lifting the entire rope 60 meters. The rope weights $60 \times 0.066 \times 9.8 = 38.808$ N, so the work applying this force for 60 meters is $60 \times 38.808 = 2,328.48$ J. This is exactly twice the work calculated before (and we leave it to the reader to understand why.)

Example $7.5.2$: Computing work performed - applying variable force

Consider again pulling a 60 m rope up a cliff face, where the rope has a mass of 66 g/m. At what point is exactly half the work performed?

Solution

From Example $7.5.1$ we know the total work performed is $1,164.24$ J. We want to find a height h such that the work in pulling the rope from a height of $x = 0$ to a height of $x = h$ is 582.12, half the total work. Thus we want to solve the equation

$$\int_0^h 0.6468(60 - x)\,dx = 582.12$$

for h.

$$\begin{aligned} \int_0^h 0.6468(60 - x)\ dx &= 582.12 \\ \left(38.808x - 0.3234x^2\right)\Big|_0^h &= 582.12 \\ 38.808h - 0.3234h^2 &= 582.12 \\ -0.3234h^2 + 38.808h - 582.12 &= 0. \end{aligned}$$

Apply the Quadratic Formula.

$$h = 17.57 \text{ and } 102.43$$

As the rope is only 60 m long, the only sensible answer is $h = 17.57$. Thus about half the work is done pulling up the first 17.5 m the other half of the work is done pulling up the remaining 42.43 m.

Example $7.5.3$: Computing work performed: applying variable force

A box of 100 lb of sand is being pulled up at a uniform rate a distance of 50 ft over 1 minute. The sand is leaking from the box at a rate of 1 lb/s. The box itself weighs 5 lb and is pulled by a rope weighing .2 lb/ft.

a. How much work is done lifting just the rope?
b. How much work is done lifting just the box and sand?
c. What is the total amount of work performed?

Solution

a. We start by forming the force function $F_r(x)$ for the rope (where the subscript denotes we are considering the rope). As in the previous example, let x denote the amount of rope, in feet, pulled in. (This is the same as saying x denotes the height of the box.) The weight of the rope with x feet pulled in is

$$F_r(x) = 0.2(50 - x) = 10 - 0.2x.$$

(Note that we do not have to include the acceleration of gravity here, for the *weight* of the rope per foot is given, not its *mass* per meter as before.) The work performed lifting the rope is

$$W_r = \int_0^{50} (10 - 0.2x)\ dx = 250 \text{ ft-lb.}$$

b. The sand is leaving the box at a rate of 1 lb/s. As the vertical trip is to take one minute, we know that 60 lb will have left when the box reaches its final height of 50 ft. Again letting x represent the height of the box, we have two points on the line that describes the weight of the sand: when $x = 0$, the sand weight is 100 lb, producing the point $(0, 100)$; when $x = 50$, the sand in the box weighs 40 lb, producing the point $(50, 40)$. The slope of this line is $\frac{100-40}{0-50} = -1.2$, giving the equation of the weight of the sand at height x as $w(x) = -1.2x + 100$. The box itself weighs a constant 5 lb, so the total force function is $F_b(x) = -1.2x + 105$. Integrating from $x = 0$ to $x = 50$ gives the work performed in lifting box and sand:

$$W_b = \int_0^{50} (-1.2x + 105)\ dx = 3750 \text{ ft-lb.}$$

c. The total work is the sum of W_r and W_b: $250 + 3750 = 4000$ ft-lb. We can also arrive at this via integration:

$$\begin{aligned} W &= \int_0^{50} (F_r(x) + F_b(x))\, dx \\ &= \int_0^{50} (10 - 0.2x - 1.2x + 105)\, dx \\ &= \int_0^{50} (-1.4x + 115)\, dx \\ &= 4000 \text{ ft-lb.} \end{aligned}$$

Hooke's Law and Springs

Hooke's Law states that the force required to compress or stretch a spring x units from its natural length is proportional to x; that is, this force is $F(x) = kx$ for some constant k. For example, if a force of 1 N stretches a given spring 2 cm, then a force of 5 N will stretch the spring 10 cm. Converting the distances to meters, we have that stretching this spring 0.02 m requires a force of $F(0.02) = k(0.02) = 1$ N, hence $k = \frac{1}{0.02} = 50$ N/m.

Example $7.5.4$: Computing work performed: stretching a spring

A force of 20 lb stretches a spring from a natural length of 7 inches to a length of 12 inches. How much work was performed in stretching the spring to this length?

Solution

In many ways, we are not at all concerned with the actual length of the spring, only with the amount of its change. Hence, we do not care that 20 lb of force stretches the spring to a length of 12 inches, but rather that a force of 20 lb stretches the spring by 5 in. This is illustrated in Figure $7.5.1$; we only measure the change in the spring's length, not the overall length of the spring.

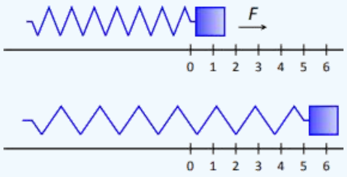


Figure $7.5.1$: Illustrating the important aspects of stretching a spring in computing work in Example $7.5.4$.

Converting the units of length to feet, we have

$$F\left(\frac{5}{12}\right) = \frac{5}{12}k = 20 \text{ lb.}$$

Thus $k = 48$ lb/ft and $F(x) = 48x$.

We compute the total work performed by integrating $F(x)$ from $x = 0$ to $x = \frac{5}{12}$:

$$\begin{aligned} W &= \int_0^{5/12} 48x\,dx \\ &= 24x^2\Big|_0^{5/12} \\ &= \frac{25}{6} \approx 4.1667 \text{ ft-lb.} \end{aligned}$$

Pumping Fluids

Another useful example of the application of integration to compute work comes in the pumping of fluids, often illustrated in the context of emptying a storage tank by pumping the fluid out the top. This situation is different than our previous examples for the forces involved are constant. After all, the force required to move one cubic foot of water (about 62.4 lb) is the same regardless of its location in the tank. What is variable is the distance that cubic foot of water has to travel; water closer to the top travels less distance than water at the bottom, producing less work.

Figure $7.5.2$: Weight and Mass densities.

Fluid	lb/ft^3	kg/m^3
Concrete	150	2400
Fuel Oil	55.46	890.13
Gasoline	45.93	737.22
Iodine	307	4927
Methanol	49.3	791.3
Mercury	844	1354
Milk	63.6-65.4	1020-1050
Water	62.4	1000

We demonstrate how to compute the total work done in pumping a fluid out of the top of a tank in the next two examples.

Example $7.5.5$: Computing work performed: pumping fluids

A cylindrical storage tank with a radius of 10 ft and a height of 30 ft is filled with water, which weighs approximately 62.4 lb/ft^3. Compute the amount of work performed by pumping the water up to a point 5 feet above the top of the tank.

Solution

We will refer often to Figure $7.5.3$ which illustrates the salient aspects of this problem.

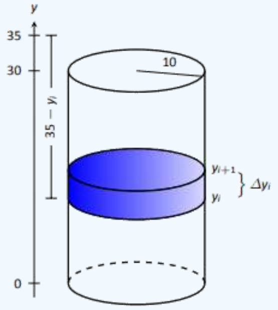


Figure $7.5.3$: Illustrating a water tank in order to compute the work required to empty it in Example 7.5.5.

We start as we often do: we partition an interval into subintervals. We orient our tank vertically since this makes intuitive sense with the base of the tank at $y = 0$. Hence the top of the water is at $y = 30$, meaning we are interested in subdividing the y-interval $[0, 30]$ into n subintervals as

$$0 = y_1 < y_2 < \cdots < y_{n+1} = 30.$$

Consider the work W_i of pumping only the water residing in the i^{th} subinterval, illustrated in Figure $7.5.3$. The force required to move this water is equal to its weight which we calculate as volume $\times$ density. The volume of water in this subinterval is $V_i = 10^2\pi\Delta y_i$; its density is 62.4 lb/ft³. Thus the required force is $6240\pi\Delta y_i$ lb.

We approximate the distance the force is applied by using any y-value contained in the i^{th} subinterval; for simplicity, we arbitrarily use y_i for now (it will not matter later on). The water will be pumped to a point 5 feet above the top of the tank, that is, to the height of $y = 35$ ft. Thus the distance the water at height y_i travels is $35 - y_i$ ft.

In all, the approximate work W_i performed in moving the water in the i^{th} subinterval to a point 5 feet above the tank is

$$W_i \approx 6240\pi\Delta y_i(35 - y_i).$$

To approximate the total work performed in pumping out all the water from the tank, we sum all the work W_i performed in pumping the water from each of the n subintervals of $[0, 30]$:

$$W \approx \sum_{i=1}^{n} W_i = \sum_{i=1}^{n} 6240\pi\Delta y_i(35 - y_i).$$

This is a Riemann sum. Taking the limit as the subinterval length goes to 0 gives

$$\begin{aligned} W &= \int_0^{30} 6240\pi(35 - y)\,dy \\ &= 6240\pi\left(35y - \tfrac{1}{2}y^2\right)\Big|_0^{30} \\ &= 11,762,123 \text{ ft-lb} \\ &\approx 1.176 \times 10^7 \text{ ft-lb}. \end{aligned}$$

We can "streamline" the above process a bit as we may now recognize what the important features of the problem are. Figure $7.5.4$ shows the tank from Example $7.5.5$ without the i^{th} subinterval identified.

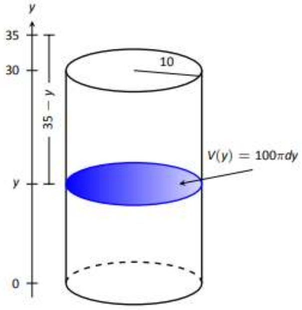


Figure $7.5.4$: A simplified illustration for computing work.

Instead, we just draw one differential element. This helps establish the height a small amount of water must travel along with the force required to move it (where the force is volume $\times$ density).

We demonstrate the concepts again in the next examples.

Example $7.5.6$: Computing work performed - pumping fluids

A conical water tank has its top at ground level and its base 10 feet below ground. The radius of the cone at ground level is 2 ft. It is filled with water weighing 62.4 lb/ft³ and is to be emptied by pumping the water to a spigot 3 feet above ground level. Find the total amount of work performed in emptying the tank.

Solution

The conical tank is sketched in Figure $7.5.5$. We can orient the tank in a variety of ways; we could let $y = 0$ represent the base of the tank and $y = 10$ represent the top of the tank, but we choose to keep the convention of the wording given in the problem and let $y = 0$ represent ground level and hence $y = -10$ represents the bottom of the tank. The actual "height" of the water does not matter; rather, we are concerned with the distance the water travels.

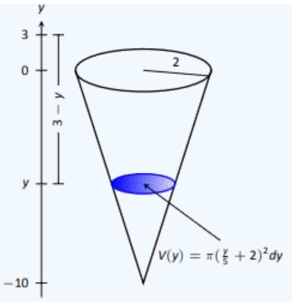


Figure 7.5.5: A graph of the conical water tank in Example 7.5.6.

The figure also sketches a differential element, a cross-sectional circle. The radius of this circle is variable, depending on y. When $y=-10$, the circle has radius 0; when $y=0$, the circle has radius 2. These two points, $(-10,0)$ and $(0,2)$, allow us to find the equation of the line that gives the radius of the cross-sectional circle, which is $r(y)=\frac{1}{5}y+2$. Hence the volume of water at this height is $V(y)=\pi(\frac{1}{5}y+2)^2\,dy$, where dy represents a very small height of the differential element. The force required to move the water at height y is $F(y)=62.4\times V(y)$.

The distance the water at height y travels is given by $h(y)=3-y$. Thus the total work done in pumping the water from the tank is

$$\begin{aligned} W &= \int_{-10}^{0} 62.4\pi(\tfrac{1}{5}y+2)^2(3-y)\,dy \\ &= 62.4\pi\int_{-10}^{0}\left(-\tfrac{1}{25}y^3-\tfrac{17}{25}y^2-\tfrac{8}{5}y+12\right)\,dy \\ &= 62.2\pi\cdot\frac{220}{3} \approx 14,376 \text{ ft-lb.} \end{aligned}$$

Example 7.5.7: Computing work performed - pumping fluids

A rectangular swimming pool is 20 ft wide and has a 3 ft "shallow end" and a 6 ft "deep end." It is to have its water pumped out to a point 2 ft above the current top of the water.

The cross-sectional dimensions of the water in the pool are given in Figure 7.5.6; note that the dimensions are for the water, not the pool itself. Compute the amount of work performed in draining the pool.

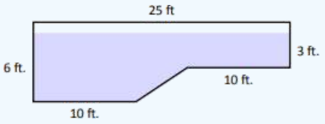


Figure 7.5.6: The cross-section of a swimming pool filled with water in Example 7.5.7.

Solution

For the purposes of this problem we choose to set $y=0$ to represent the bottom of the pool, meaning the top of the water is at $y = 6$.

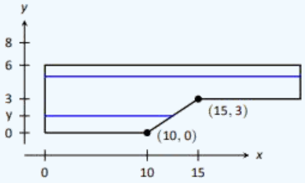


Figure 7.5.7: Orienting the pool and showing differential elements for Example 7.5.7.

Figure 7.5.7 shows the pool oriented with this y-axis, along with 2 differential elements as the pool must be split into two different regions.

The top region lies in the y-interval of $[3,6]$, where the length of the differential element is 25 ft as shown. As the pool is 20 ft wide, this differential element represents a this slice of water with volume $V(y)=20\cdot 25\cdot dy$. The water is to be pumped to a height of $y=8$, so the height function is $h(y)=8-y$. The work done in pumping this top region of water is

$$W = 62.4\int^{6} 500(8-y)\,dy = 327,600 \text{ ft-lb}$$

$$W_t = 62.4\int_3 500(8-y)\,dy = 327,600\text{ ft-lb.}$$

The bottom region lies in the y-interval of $[0,3]$; we need to compute the length of the differential element in this interval.

One end of the differential element is at $x=0$ and the other is along the line segment joining the points $(10,0)$ and $(15,3)$. The equation of this line is $y=\frac{3}{5}(x-10)$; as we will be integrating with respect to y, we rewrite this equation as $x=\frac{5}{3}y+10$. So the length of the differential element is a difference of x-values: $x=0$ and $x=\frac{5}{3}y+10$, giving a length of $x=\frac{5}{3}y+10$.

Again, as the pool is 20 ft wide, this differential element represents a thin slice of water with volume $V(y)=20\cdot(\frac{5}{3}y+10)\cdot dy$; the height function is the same as before at $h(y)=8-y$. The work performed in emptying this part of the pool is

$$W_b = 62.4\int_0^3 20(\tfrac{5}{3}y+10)(8-y)\,dy = 299,520\text{ ft-lb.}$$

The total work in emptying the pool is

$$W = W_b + W_t = 327,600 + 299,520 = 627,120\text{ ft-lb.}$$

Notice how the emptying of the bottom of the pool performs almost as much work as emptying the top. The top portion travels a shorter distance but has more water. In the end, this extra water produces more work.

The next section introduces one final application of the definite integral, the calculation of fluid force on a plate.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 7.5: Work is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

7.6: Fluid Forces

In the unfortunate situation of a car driving into a body of water, the conventional wisdom is that the water pressure on the doors will quickly be so great that they will be effectively unopenable. (Survival techniques suggest immediately opening the door, rolling down or breaking the window, or waiting until the water fills up the interior at which point the pressure is equalized and the door will open. See Mythbusters episode #72 to watch Adam Savage test these options.) How can this be true? How much force does it take to open the door of a submerged car? In this section we will find the answer to this question by examining the forces exerted by fluids.

We start with **pressure**, which is related to **force** by the following equations:

$$\text{Pressure} = \frac{\text{Force}}{\text{Area}} \quad \Leftrightarrow \quad \text{Force} = \text{Pressure} \times \text{Area}. \tag{7.6.1}$$

In the context of fluids, we have the following definition.

Definition 26: Fluid pressure

Let w be the weight-density of a fluid. The *pressure* p exerted on an object at depth d in the fluid is $p = w \cdot d$.

We use this definition to find the **force** exerted on a horizontal sheet by considering the sheet's area.

Example $7.6.1$: Computing fluid force

1. A cylindrical storage tank has a radius of 2 ft and holds 10 ft of a fluid with a weight-density of 50 lb/ft³ (Figure $7.6.1$.) What is the force exerted on the base of the cylinder by the fluid?

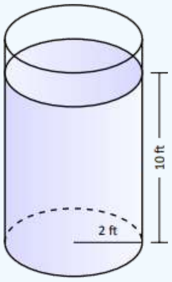


Figure $7.6.1$: A cylindrical tank in Example $7.6.1$.

1. A rectangular tank whose base is a 5 ft square has a circular hatch at the bottom with a radius of 2 ft. The tank holds 10 ft of a fluid with a weight-density of 50 lb/ft³. (Figure $7.6.2$). What is the force exerted on the hatch by the fluid?

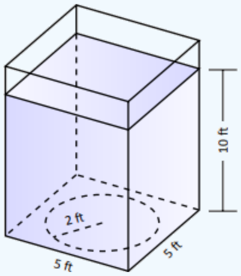


Figure $7.6.2$: A rectangular tank in Example $7.6.1$.

Solution

1. Using Definition 26, we calculate that the pressure exerted on the cylinder's base is $w \cdot d = 50 \text{ lb/ft}^3 \times 10 \text{ ft} = 500$ lb/ft². The area of the base is $\pi \cdot 2^2 = 4\pi$ ft². So the force exerted by the fluid is

$$F = 500 \times 4\pi = 6283 \text{ lb}.$$

Note that we effectively just computed the *weight* of the fluid in the tank.

2. The dimensions of the tank in this problem are irrelevant. All we are concerned with are the dimensions of the hatch and the depth of the fluid. Since the dimensions of the hatch are the same as the base of the tank in the previous part of this example, as is the depth, we see that the fluid force is the same. That is, $F = 6283$ lb.

A key concept to understand here is that we are effectively measuring the weight of a 10 ft column of water above the hatch. The size of the tank holding the fluid does not matter.

The previous example demonstrates that computing the force exerted on a horizontally oriented plate is relatively easy to compute. What about a vertically oriented plate? For instance, suppose we have a circular porthole located on the side of a submarine. How do we compute the fluid force exerted on it?

Pascal's Principle states that the pressure exerted by a fluid at a depth is equal in all directions. Thus the pressure on any portion of a plate that is 1 ft below the surface of water is the same no matter how the plate is oriented. (Thus a hollow cube submerged at a great depth will not simply be "crushed" from above, but the sides will also crumple in. The fluid will exert force on *all* sides of the cube.)

So consider a vertically oriented plate as shown in Figure $7.6.3$ submerged in a fluid with weight-density w. What is the total fluid force exerted on this plate? We find this force by first approximating the force on small horizontal strips.

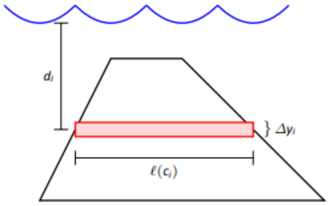


Figure $7.6.3$: A thin, vertically oriented plate submerged in a fluid with weight density w.

Let the top of the plate be at depth b and let the bottom be at depth a. (For now we assume that surface of the fluid is at depth 0, so if the bottom of the plate is 3 ft under the surface, we have $a = -3$. We will come back to this later.) We partition the interval $[a, b]$ into n subintervals

$$a = y_1 < y_2 < \cdots < y_{n+1} = b, \tag{7.6.2}$$

with the $i^{\text{ th}}$ subinterval having length Δy_i. The force F_i exerted on the plate in the $i^{\text{ th}}$ subinterval is $F_i = \text{Pressure} \times \text{Area}$.

The pressure is depth $\times w$. We approximate the depth of this thin strip by choosing any value d_i in $[y_i, y_{i+1}]$; the depth is approximately $-d_i$. (Our convention has d_i being a negative number, so $-d_i$ is positive.) For convenience, we let d_i be an endpoint of the subinterval; we let $d_i = y_i$.

The area of the thin strip is approximately length $\times$ width. The width is Δy_i. The length is a function of some y-value c_i in the $i^{\text{ th}}$ subinterval. We state the length is $\ell(c_i)$. Thus

$$\begin{aligned} F_i &= \text{Pressure} \times \text{Area} \\ &= -y_i \cdot w \times \ell(c_i) \cdot \Delta y_i. \end{aligned}$$

To approximate the total force, we add up the approximate forces on each of the n thin strips:

$$F = \sum_{i=1}^{n} F_i \approx \sum_{i=1}^{n} -w \cdot y_i \cdot \ell(c_i) \cdot \Delta y_i. \tag{7.6.3}$$

This is, of course, another Riemann Sum. We can find the exact force by taking a limit as the subinterval lengths go to 0; we evaluate this limit with a definite integral.

key idea 30: fluid force on a vertically oriented plate

Let a vertically oriented plate be submerged in a fluid with weight-density w where the top of the plate is at $y = b$ and the bottom is at $y = a$. Let $\ell(y)$ be the length of the plate at y.

1. If $y = 0$ corresponds to the surface of the fluid, then the force exerted on the plate by the fluid is

$$F = \int^{b} w \cdot (-y) \cdot \ell(y) dy. \tag{7.6.4}$$

2. In general, let $d(y)$ represent the distance between the surface of the fluid and the plate at y. Then the force exerted on the plate by the fluid is

$$F = \int_a^b w \cdot d(y) \cdot \ell(y)\ dy. \tag{7.6.5}$$

Example $7.6.2$: Finding fluid force

Consider a thin plate in the shape of an isosceles triangle as shown in Figure $7.6.4$ submerged in water with a weight-density of 62.4 lb/ft^3. If the bottom of the plate is 10 ft below the surface of the water, what is the total fluid force exerted on this plate?

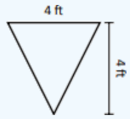


Figure $7.6.4$: A thin plate in the shape of an isosceles triangle in Example 7.6.2.

Solution

We approach this problem in two different ways to illustrate the different ways Key Idea 30 can be implemented. First we will let $y = 0$ represent the surface of the water, then we will consider an alternate convention.

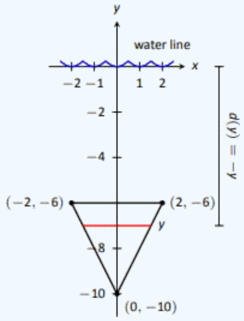


Figure $7.6.5$: Sketching the triangular plate in Example 7.6.2 with the convention that the water level is at y=0.

1. We let $y = 0$ represent the surface of the water; therefore the bottom of the plate is at $y = -10$. We center the triangle on the y-axis as shown in Figure $7.6.5$. The depth of the plate at y is $-y$ as indicated by the Key Idea. We now consider the length of the plate at y.

   We need to find equations of the left and right edges of the plate. The right hand side is a line that connects the points $(0, -10)$ and $(2, -6)$: that line has equation $x = 1/2(y + 10)$. (Find the equation in the familiar $y = mx + b$ format and solve for x.) Likewise, the left hand side is described by the line $x = -1/2(y + 10)$. The total length is the distance between these two lines: $\ell(y) = 1/2(y + 10) - (-1/2(y + 10)) = y + 10$.

   The total fluid force is then

$$\begin{aligned} F &= \int_{-10}^{-6} 62.4(-y)(y+10)dy \\ &= 62.4 \cdot \frac{176}{3} \approx 3660.8 \text{ lb.} \end{aligned}$$

2. Sometimes it seems easier to orient the thin plate nearer the origin. For instance, consider the convention that the bottom of the triangular plate is at $(0, 0)$, as shown in Figure $7.6.6$. The equations of the left and right hand sides are easy to find. They are $y = 2x$ and $y = -2x$, respectively, which we rewrite as $x = 1/2y$ and $x = -1/2y$. Thus the length function is $\ell(y) = 1/2y - (-1/2y) = y$.

   As the surface of the water is 10 ft above the base of the plate, we have that the surface of the water is at $y = 10$. Thus the depth function is the distance between $y = 10$ and y; $d(y) = 10 - y$. We compute the total fluid force as:

$$
\begin{aligned}
F &= \int_0^4 62.4(10-y)(y)\ dy \\
&\approx 3660.8 \text{ lb.}
\end{aligned}
$$

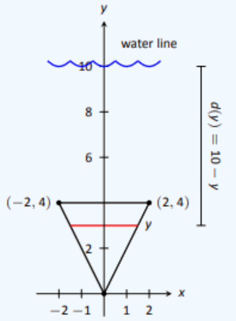


Figure 7.6.6: Sketching the triangular plate in Example 7.6.2 with the convention that the base of the triangle is at $(0,0)$.

The correct answer is, of course, independent of the placement of the plate in the coordinate plane as long as we are consistent.

Example 7.6.3: Finding fluid force

Find the total fluid force on a car door submerged up to the bottom of its window in water, where the car door is a rectangle 40" long and 27" high (based on the dimensions of a 2005 Fiat Grande Punto.)

Solution

The car door, as a rectangle, is drawn in Figure 7.6.7. Its length is $10/3$ ft and its height is 2.25 ft. We adopt the convention that the top of the door is at the surface of the water, both of which are at $y=0$. Using the weight-density of water of 62.4 lb/ft^3, we have the total force as

$$
\begin{aligned}
F &= \int_{-2.25}^{0} 62.4(-y)10/3dy \\
&= \int_{-2.25}^{0} -208ydy \\
&= -104y^2\Big|_{-2.25}^{0} \\
&= 526.5 \text{ lb.}
\end{aligned}
$$

Most adults would find it very difficult to apply over 500 lb of force to a car door while seated inside, making the door effectively impossible to open. This is counter-intuitive as most assume that the door would be relatively easy to open. The truth is that it is not, hence the survival tips mentioned at the beginning of this section.

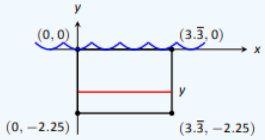


Figure 7.6.7: Sketching a submerged car door in Example 7.6.2.

Example 7.6.4: Finding fluid force

An underwater observation tower is being built with circular viewing portholes enabling visitors to see underwater life. Each vertically oriented porthole is to have a 3 ft diameter whose center is to be located 50 ft underwater. Find the total fluid force exerted on each porthole. Also, compute the fluid force on a horizontally oriented porthole that is under 50 ft of water.

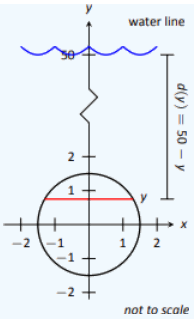


Figure 7.6.8: Measuring the fluid force on an underwater porthole in Example 7.6.4.

Solution

We place the center of the porthole at the origin, meaning the surface of the water is at $y = 50$ and the depth function will be $d(y) = 50 - y$; see Figure 7.6.8.

The equation of a circle with a radius of 1.5 is $x^2 + y^2 = 2.25$; solving for x we have $x = \pm\sqrt{2.25 - y^2}$, where the positive square root corresponds to the right side of the circle and the negative square root corresponds to the left side of the circle. Thus the length function at depth y is $\ell(y) = 2\sqrt{2.25 - y^2}$. Integrating on $[-1.5, 1.5]$ we have:

$$\begin{aligned} F &= 62.4\int_{-1.5}^{1.5} 2(50-y)\sqrt{2.25-y^2}dy \\ &= 62.4\int_{-1.5}^{1.5}\left(100\sqrt{2.25-y^2} - 2y\sqrt{2.25-y^2}\right)dy \\ &= 6240\int_{-1.5}^{1.5}\left(\sqrt{2.25-y^2}\right)dy - 62.4\int_{-1.5}^{1.5}\left(2y\sqrt{2.25-y^2}\right)dy. \end{aligned}$$

The second integral above can be evaluated using Substitution. Let $u = 2.25 - y^2$ with $du = -2y\,dy$. The new bounds are: $u(-1.5) = 0$ and $u(1.5) = 0$; the new integral will integrate from $u = 0$ to $u = 0$, hence the integral is 0.

The first integral above finds the area of half a circle of radius 1.5, thus the first integral evaluates to $6240 \cdot \pi \cdot 1.5^2/2 = 22,054$. Thus the total fluid force on a vertically oriented porthole is $22,054$ lb.

Finding the force on a horizontally oriented porthole is more straightforward:

$$F = \text{Pressure} \times \text{Area} = 62.4 \cdot 50 \times \pi \cdot 1.5^2 = 22,054 \text{ lb}. \tag{7.6.6}$$

That these two forces are equal is not coincidental; it turns out that the fluid force applied to a vertically oriented circle whose center is at depth d is the same as force applied to a horizontally oriented circle at depth d.

We end this chapter with a reminder of the true skills meant to be developed here. We are not truly concerned with an ability to find fluid forces or the volumes of solids of revolution. Work done by a variable force is important, though measuring the work done in pulling a rope up a cliff is probably not.

What we are actually concerned with is the ability to solve certain problems by first approximating the solution, then refining the approximation, then recognizing if/when this refining process results in a definite integral through a limit. Knowing the formulas found inside the special boxes within this chapter is beneficial as it helps solve problems found in the exercises, and other mathematical skills are strengthened by properly applying these formulas. However, more importantly, understand how each of these formulas was constructed. Each is the result of a summation of approximations; each summation was a Riemann sum, allowing us to take a limit and find the exact answer through a definite integral.

The next chapter addresses an entirely different topic: sequences and series. In short, a sequence is a list of numbers, where a series is the summation of a list of numbers. These seemingly-simple ideas lead to very powerful mathematics.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 7.6: Fluid Forces is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

7.E: Applications of Integration (Exercises)

7.1: Area Between Curves

Terms and Concepts

1. T/F: The are between curves is always positive.

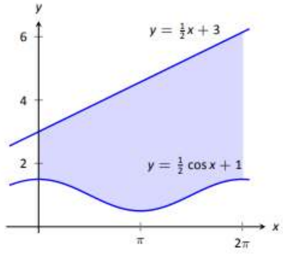
2. T/F: Calculus can be used to find the area of basic geometric shapes.

3. In your own words, describe how to find the total area enclosed by $y = f(x)$ and $y = g(x)$.

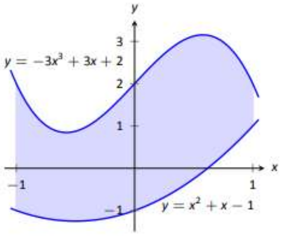
Problems

In Exercises 4-10, find the area of the shaded region in the given graph.

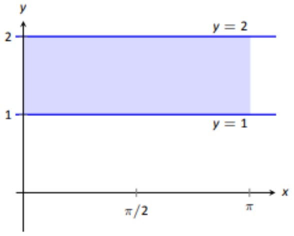
4.



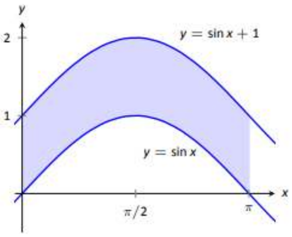
5.



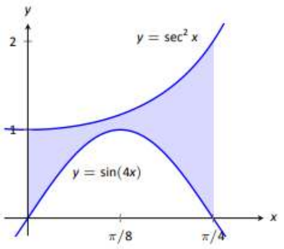
6.



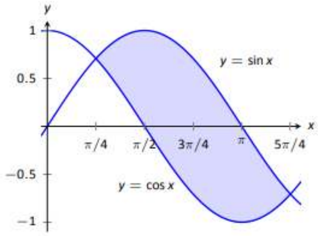
7.



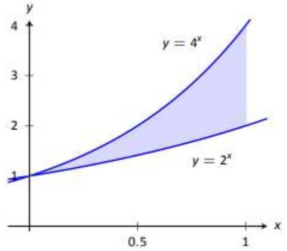
8.



9.



10.



In Exercises 11-16, find the total area enclosed by the functions f and g.

11. $f(x)=2x^2+5x-3,\ g(x)=x^2+4x-1$

12. $f(x)=x^2-3x+2,\ g(x)=-3x+3$

13. $f(x)=\sin x,\ g(x)=2x/\pi$

14.. $f(x)=x^3-4x^2+x-1,\ g(x)=-x^2+2x-4$

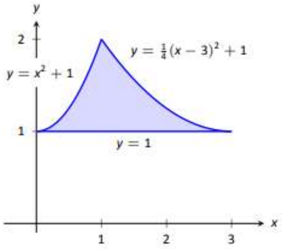
15. $f(x)=x,\ g(x)=\sqrt{x}$

16. $f(x)=-x^3+5x^2+2x+1,\ g(x)=3x^2+x+3$

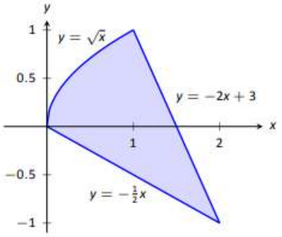
17. The functions $f(x)=\cos(2x)$ and $g(x)=\sin x$ intersect infinitely many times, forming an infinite number of repeated, enclosed regions. Find the areas of these regions.

In Exercises 18-22, find the area of the enclosed region in two ways:
1. by treating the boundaries as functions of x, and
2. by treating the boundaries as functions of y.

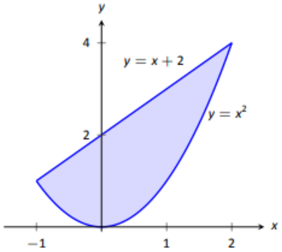
18.



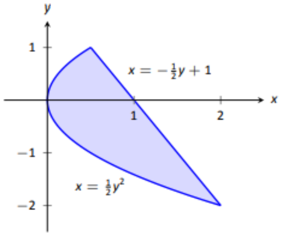
19.



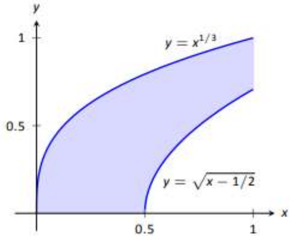
20.



21.



22.



In Exercises 23-26, find the are triangle formed by the given three points.

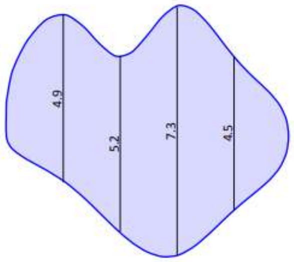
23. $(1,1)$, $(2,3)$ and $(3,3)$

24. $(-1,1)$, $(1,3)$ and $(2,-1)$

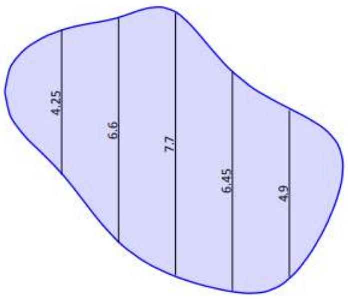
25. $(1,1)$, $(3,3)$ and $(3,3)$

26. $(0,0)$, $(2,5)$ and $(5,2)$

27. Use the Trapezoidal Rule to approximate the area of the pictured lake whose lengths, in hundreds of feet, are measured in 100-foot increments.



28. Use Simpson's Rule to approximate the area of the pictured lake whose lengths, in hundreds of feet, are measured in 200-foot increments.



7.2: Volume by Cross-Sectional Area: Disk and Washer Methods

Terms and Concepts

1. T/F: A solid of revolution is formed by revolving a shape around an axis.

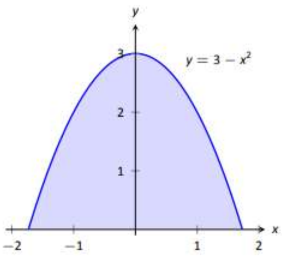
2. In your own words, explain how the Disk and Washer Methods are related.

3 Explain how the units of volume are found in the integral of Theorem 54: if $A(x)$ has units of in^2, how does $\int A(x)\,dx$ have units of in^3?

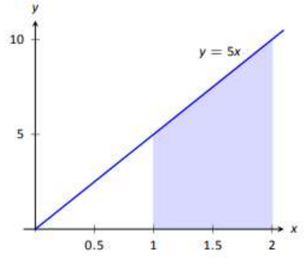
Problems

In Exercises 4-7, a region of the Cartesian plane is shaded. Use the Disk/Washer Method to find the volume of the solid of revolution formed by revolving the region about the x-axis.

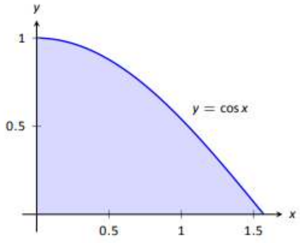
4.



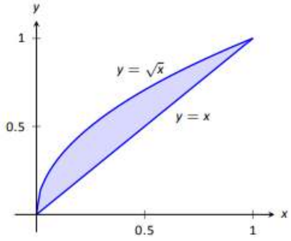
5.



6.

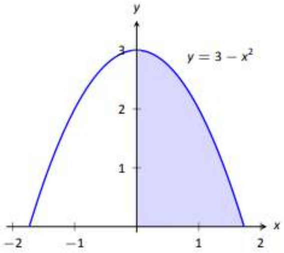


7.

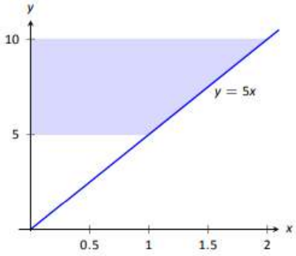


In Exercises 8-11, a region of the Cartesian plane is shaded. Use the Disk/Washer Method to find the volume of the solid of revolution formed by revolving the region about the y-axis.

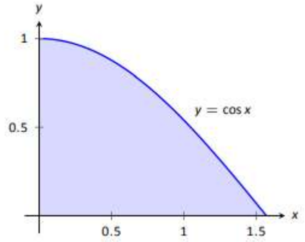
8.



9.

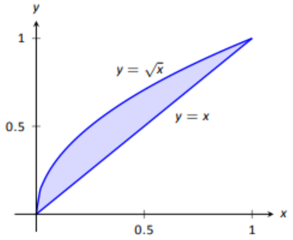


10.



(Hint: Integration By Parts will be necessary, twice. First let $u = \arccos^2 x$, then let $u = \arccos x$.)

11.



In Exercises 12-17, a region of the Cartesian plane is described. Use the Disk/Washer Method to find the volume of the solid of revolution formed by rotating the region about each of the given axes.

12. Region bounded by: $y = \sqrt{x}$, $y = 0$ and $x = 1$.
Rotate about:
(a) the *x*-axis
(b) $y = 1$
(c) the *y*-axis
(d) $x = 1$

13. Region bounded by: $y = 4 - x^2$ and $y = 0$.
Rotate about:
(a) the *x*-axis
(b) $y = 4$
(c) $y = -1$
(d) $x = 2$

14. The triangle with vertices $(1, 1)$, $(1, 2)$ and $(2, 1)$.
Rotate about:
(a) the *x*-axis
(b) $y = 2$
(c) the *y*-axis
(d) $x = 1$

15. Region bounded by: $y = y = x^2 - 2x + 2,$ and $y = 2x - 1$.
Rotate about:
(a) the *x*-axis
(b) $y = 1$
(c) $y = 5$

16. Region bounded by: $y = 1/\sqrt{x^2 + 1}$, $x = -1$, $x = 1$ and the x-axis.
Rotate about:
(a) the *x*-axis
(b) $y = 1$
(c) $y = -1$

17. Region bounded by: $y = 2x,\ y = x$ and $x = 2.$

Rotate about:

(a) the *x*-axis

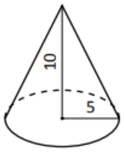
(b) $y = 4$

(c) the *y*-axis

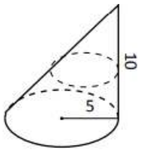
(d) $x = 2$

In Exercises 18-21, a solid is described. Orient the solid along the x-axis such that a cross-sectional area function $A(x)$ can be obtained, then apply Theorem 54 to find the volume of the solid.

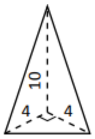
18. A right circular cone with height of 10 and base radius of 5.



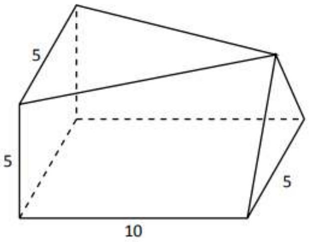
19. A skew right circular cone with height of 10 and base radius of 5. (Hint: all cross-sections are circles.)



20. A right triangle cone with height of 10 and whose base is a right, isosceles triangle with side length 4.



21. A solid with length 10 with a rectangular base and triangular top, wherein one end is a square with side length 5 and the other end is a triangle with base and height of 5.



7.3: The Shell Method

Terms and Concepts

1. T/F: A solid of revolution is formed by revolving a shape around an axis.

2. T/F: The Shell Method can only be used when the Washer Method fails.

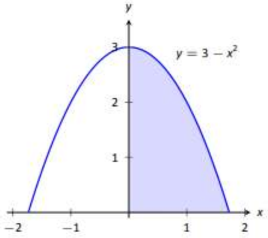
3. T/F: The Shell Method works by integrating cross-sectional areas of a solid.

4. T/F: When finding the volume of a solid of revolution that was revolved around a vertical axis, the Shell Method integrates with respect to *x*.

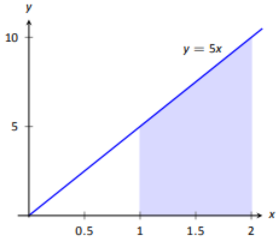
Problems

In Exercises 5-8, a region of the Cartesian plane is shaded. Use the Shell Method to find the volume of the solid of revolution formed by revolving the region about the y-axis.

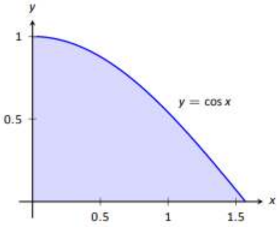
5.



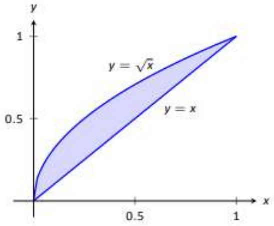
6.



7.

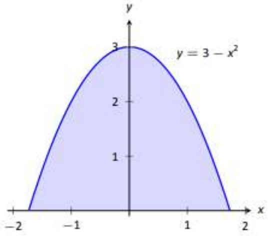


8.

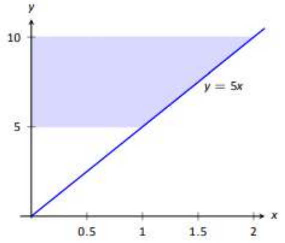


In Exercises 9-12, a region of the Cartesian plane is shaded. Use the Shell Method to find the volume of the solid of revolution formed by revolving the region about the x-axis.

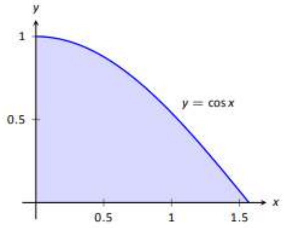
9.



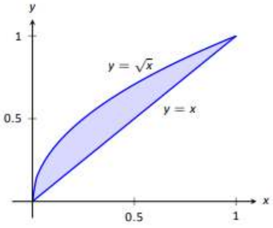
10.



11.



12.



In Exercises 13-18, a region of the Cartesian plane is described. Use the Shell Method to find the volume of the solid of revolution formed by rotating the region about each of the given axes.

13. Region bounded by: $y=\sqrt{x},\ y=0 \text{ and } x=1.$
Rotate about:
(a) the *x*-axis
(b) $y=1$
(c) the *y*-axis
(d) $x=1$

14. Region bounded by: $y=4-x^2 \text{ and } y=0.$
Rotate about:
(a) $x=2$
(b) $x=-2$
(c) the x-axis
(d) $y=4$

15. The triangle with vertices $(1,1),\ (1,2) \text{ and } (2,1).$
Rotate about:
(a) the *x*-axis
(b) $x=1$
(c) the *x*-axis
(d) $y=2$

16. Region bounded by: $y=y=x^2-2x+2,\ \text{ and } y=2x-1.$
Rotate about:
(a) the *x*-axis

(b) $x=1$
(c) $x=-1$

17. Region bounded by: $y=1/\sqrt{x^2+1},\ x=1$ and the x and y-axis.
Rotate about:
(a) the *x*-axis
(b) $x=1$

18. Region bounded by: $y=2x,\ y=x$ and $x=2$.
Rotate about:
(a) the *y*-axis
(b) $x=2$
(c) the *x*-axis
(d) $y=4$

7.4: Arc Length and Surface Area

Terms and Concepts

1. T/F: The integral formula for computing Arc Length was found by first approximating arc length with straight line segments.

2. T/F: The integral formula for computing Arc Length includes a square-root, meaning the integration is probably easy.

Problems

In Exercises 3-12, find the arc length of the function on the given interval.

3. $f(x)=x$ on $[0,1]$.

4. $f(x)=\sqrt{8}x$ on $[-1,1]$.

5. $f(x)=\frac{1}{3}x^{3/2}-x^{1/2}$ on $[0,1]$.

6. $f(x)=\frac{1}{12}x^3+\frac{1}{x}$ on $[1,4]$.

7. $f(x)=2x^{3/2}-\frac{1}{6}\sqrt{x}$ on $[0,9]$.

8. $f(x)=\cosh x$ on $[-\ln 2,\ln 2]$.

9. $f(x)=\frac{1}{2}(e^2+e^{-x})$ on $[0,\ln 5]$.

10. $f(x)=\frac{1}{12}x^5+\frac{1}{5x^3}$ on $[0.1,1]$.

11. $f(x)=\ln(\sin x)$ on $[\pi/6,\pi/2]$.

12. $f(x)=\ln(\cos x)$ on $[0,\pi/4]$.

In Exercises 13-20, set up the integral to compute the arc length of the function on the given interval. Do no evaluate the integral.

13. $f(x)=x^2$ on $[0,1]$.

14. $f(x)=x^{10}$ on $[0,1]$.

15. $f(x)=\sqrt{x}$ on $[0,1]$.

16. $f(x)=\ln x$ on $[1,e]$.

17. $f(x)=\sqrt{1-x^2}$ on $[-1,1]$.(Note: this describes the top half of a circle with radius 1.)

18. $f(x)=\sqrt{1-x^2/9}$ on $[-3,3]$. (Note: this describes the top half of an eclipse with a major axis of length 6 and a minor axis of length 2.)

19. $f(x)=\frac{1}{x}$ on $[1,2]$.

20. $f(x)=\sec x$ on $[-\pi/4,\pi/4]$.

In Exercises 21-28, use Simpson's Rule, with $n=4$, to approximate the arc length of the function on the given interval. Note: these are the same problems as in Exercises 13-20.

21. $f(x)=x^2$ on $[0,1]$.

22. $f(x)=x^{10}$ on $[0,1]$.

23. $f(x)=\sqrt{x}$ on $[0,1]$. (Note: $f'(x)$ is not defined at $x=0$.)

24. $f(x)=\ln x$ on $[1,e]$.

25. $f(x)=\sqrt{1-x^2}$ on $[-1,1]$. (Note: $f'(x)$ is not defined at the endpoints.)

26. $f(x)=\sqrt{1-x^2/9}$ on $[-3,3]$. (Note: $f'(x)$ is not defined at the endpoints.)

27. $f(x)=\frac{1}{x}$ on $[1,2]$.

28. $f(x)=\sec x$ on $[-\pi/4,\pi/4]$.

In Exercises 29-33, find the surface area of the described solid of revolution.

29. The solid formed by revolving $y=2x$ on $[0,1]$ about the x-axis.

30. The solid formed by revolving $y=x^2$ on $[0,1]$ about the y-axis.

31. The solid formed by revolving $y=x^3$ on $[0,1]$ about the x-axis.

32. The solid formed by revolving $y=\sqrt{x}$ on $[0,1]$ about the x-axis.

33. The solid formed by revolving $y=\sqrt{1-x^2}$ on $[-1,1]$ about the x-axis.

7.5: Work

Terms and Concepts

1. What are the typical units of work?

2. If a man has a mass of 80kg on Earth, will his mass on the moon be bigger, smaller, or the same?

3. If a woman weights 130 lb on Earth, will her weight on the moon be bigger, smaller, or the same?

Problems

4. A 100 ft rope, weighing 0.1 lb/ft, hangs over the edge of a tall building.
(a) How much work is done pulling the entire rope to the top of the building?
(b) How much rope is pulled in when half of the total work is done?

5. A 50 m rope, with a mass density of 0.2 kg/m, hangs over the edge of a tall building.
(a) How much work is done pulling the entire rope to the top of the building?
(b) How much work is done pulling in the first 20 m?

6. A rope of length l ft hangs over the edge of tall cliff. (Assume the cliff is taller than the length of the rope.) The rope has a weight density of d lb/ft.
(a) How much work is done pulling the entire rope to the top of the cliff?
(b) What percentage of the total work is done pulling in the first half of the rope?
(c) How much rope is pulled in when half of the total work is done?

7. A 20 m rope with mass density of 0.5 kg/m hangs over the edge of a 10 m building. How much work is done pulling the rope to the top?

8. A crane lifts a 2000 lb load vertically 30 ft with a 1" cable weighing 1.68 lb/ft.
(a) How much work is done lifting the cable alone?
(b) How much work is done lifting the load alone?
(c) Could one conclude that the work done lifting the cable is negligible compared to the work done lifting the load?

9. A 100 lb bag of sand is lifted uniformly 120 ft in one minute. Sand leaks from the bag at a rate of 1/4 lb/s. What is the total work done in lifting the bag?

10. A box weighing 2 lb lifts 10 lb of sand vertically 50 ft. A crack in the box allows the sand to leak out such that 9 lb of sand is in the box at the end of the trip. Assume the sand leaked out at a uniform rate. What is the total work done in lifting the box and sand?

11. A force of 1000 lb compresses a spring 3 in. How much work is performed in compressing the spring?

12. A force of 2 N stretches a spring 5 cm. How much work is performed in stretching the spring?

13. A force of 50 lb compresses a spring from a natural length of 18 in to 12 in. How much work is performed in compressing the spring?

14. A force of 20 lb stretches a spring from a natural length of 6 in to 8 in. How much work is performed in stretching the spring?

15. A force of 7 N stretches a spring from a natural length of 11 cm to 21 cm. How much work is performed in stretching the spring from a length of 16 cm to 21 cm?

16. A force of f N stretches a spring d m from its natural length. How much work is performed in stretching the spring?

17. A 20 lb weight is attached to a spring. The weight rests on the spring, compressing the spring from a natural length of 1 ft to 6 in.
How much work is done in lifting the box 1.5 ft (i.e, the spring will be stretched 1 ft beyond its natural length)?

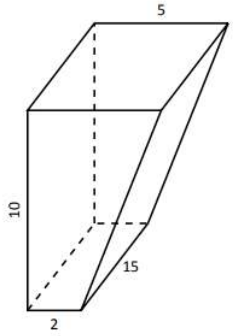
18. A 20 lb weight is attached to a spring. The weight rests on the spring, compressing the spring from a natural length of 1 ft to 6 in.
How much work is done in lifting the box 6 in (i.e, bringing the spring back to its natural length)?

19. A 5 m tall cylindrical tank with radius of 2 m is filled with 3 m of gasoline, with a mass density of 737.22 kg/m^3. Compute the total work performed in pumping all the gasoline to the top of the tank.

20. A 6 ft cylindrical tank with a radius of 3 ft is filled with water, which has a weight density of 62.4 lb/ft^3. The water is to be pumped to a point 2 ft above the top of the tank.
(a) How much work is performed in pumping all the water from the tank?
(b) How much work is performed in pumping 3 ft of water from the tank?
(c) At what point is 1/2 of the total work done?

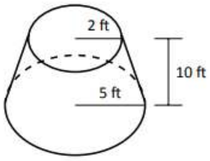
21. A gasoline tanker is filled with gasoline with a weight density of 45.93 lb/ft^3. The dispensing value at the base is jammed shut, forcing the operator to empty the tank via pumping the gas to a point 1 ft above the top of the tank. Assume the tank is a perfect cylinder, 20 ft long with a diameter of 7.5 ft. How much work is performed in pumping all the gasoline from the tank?

22. A fuel oil storage tank is 10 ft deep with trapezoidal sides, 5 ft at the top of the 2 ft at the bottom, and is 15 ft wide (see diagram below). Given that fuel oil weighs 55.46 lb/ft^3, find the work performed in pumping all the oil from the tank to a point 3 ft above the top of the tank.

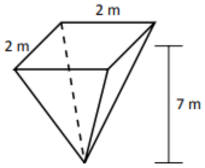


23. A conical tank is 5 m deep with a top radius of 3 m. (This is similar to Example 224.) The tank is filled with pure water, with a mass density of 1000 kg/m^3.
(a) Find the work performed in pumping all the water to the top of the tank.
(b) Find the work performed in pumping the top 2.5 m of water to the top of the tank.
(c) Find the work performed in pumping the top half of the water, by volume, to the top of the tank.

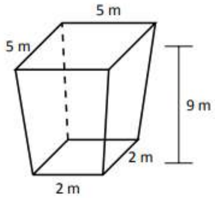
24. A water tank has the shape of a truncated cone, with dimensions given below, and is filled with water with a weight density of 62.4 lb/ft^3. Find the work performed in pumping all water to a point 1 ft above the top of the tank.



25. A water tank has the shape of an inverted pyramid, with dimensions given below, and is filled with water with a mass density of 1000 kg/m^3. Find the work performed in pumping all water to a point 5 m above the top of the tank.



26. A water tank has the shape of a truncated, inverted pyramid, with dimensions given blow, and is filled with water with a mass density of 1000 kg/m^3. Find the work performed in pumping all water to a point 1 m above the top of the tank.



7.6: Fluid Forces

Terms and Concepts

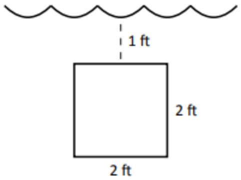
1. State in your own words Pascal's Principle.

2. State in your own words how pressure is different from force.

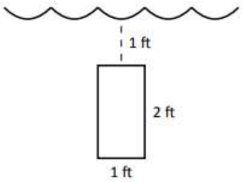
Problems

In Exercises 3-12, find the fluid force exerted on the given plate, submerged in water with a weight density of 62.4 lb/ft^3.

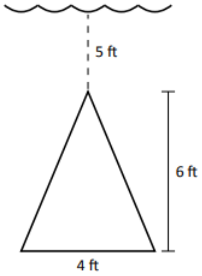
3.



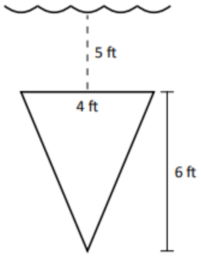
4.



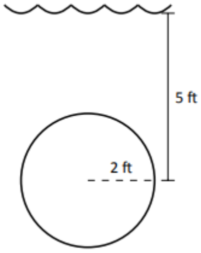
5.



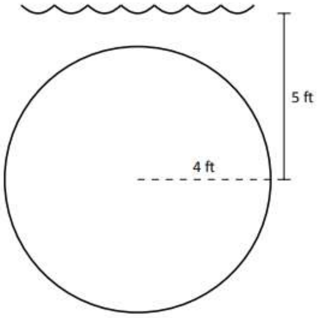
6.



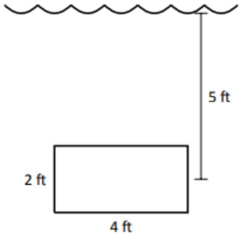
7.



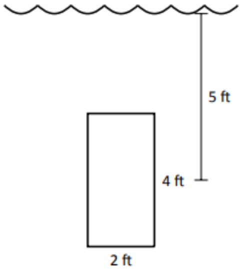
8.



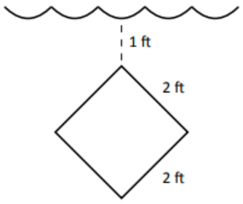
9.



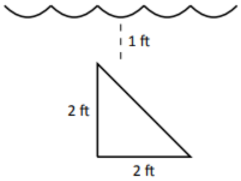
10.



11.

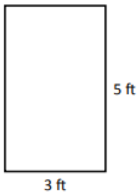


12.

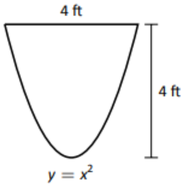


In Exercises 13-18, the side of a container is pictured. Find the fluid force exerted on this plate when the container is full of:
1. water, with a weight density of 62.4 lb/ft^3
2. concrete, with a weight density of 150 lb/ft^3.

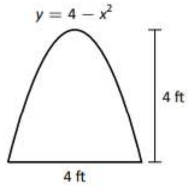
13.



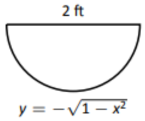
14.



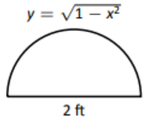
15.



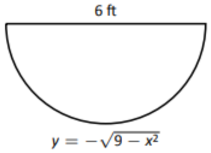
16.



17.



18.



19. How deep must the center of a vertically oriented circular plate with a radius of 1 ft be submerged in water, with a weight density of 62.4 lb/ft³, for the fluid force on the plate to reach 1,000 lb?

20. How deep must the center of a vertically oriented square plate with a side length of 2 ft be submerged in water, with a weight density of 62.4 lb/ft³, for the fluid force on the plate to reach 1,000 lb?

This page titled 7.E: Applications of Integration (Exercises) is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

8: Sequences and Series

This chapter introduces **sequences** and **series**, important mathematical constructions that are useful when solving a large variety of mathematical problems. The content of this chapter is considerably different from the content of the chapters before it. While the material we learn here definitely falls under the scope of "calculus," we will make very little use of derivatives or integrals. Limits are extremely important, though, especially limits that involve infinity.

One of the problems addressed by this chapter is this: suppose we know information about a function and its derivatives at a point, such as $f(1) = 3$, $f'(1) = 1$, $f''(1) = -2$, $f'''(1) = 7$, and so on. What can I say about $f(x)$ itself? Is there any reasonable approximation of the value of $f(2)$? The topic of Taylor Series addresses this problem, and allows us to make excellent approximations of functions when limited knowledge of the function is available.

8.1: Sequences
8.2: Infinite Series
8.3: Integral and Comparison Tests
8.4: Ratio and Root Tests
8.5: Alternating Series and Absolute Convergence
8.6: Power Series
8.7: Taylor Polynomials
8.8: Taylor Series
8.E: Applications of Sequences and Series (Exercises)

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 8: Sequences and Series is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

8.1: Sequences

We commonly refer to a set of events that occur one after the other as a *sequence* of events. In mathematics, we use the word *sequence* to refer to an ordered set of numbers, i.e., a set of numbers that "occur one after the other."

For instance, the numbers 2, 4, 6, 8, ..., form a sequence. The order is important; the first number is 2, the second is 4, etc. It seems natural to seek a formula that describes a given sequence, and often this can be done. For instance, the sequence above could be described by the function $a(n) = 2n$, for the values of $n = 1, 2, \ldots$ To find the 10^{th} term in the sequence, we would compute $a(10)$. This leads us to the following, formal definition of a sequence.

Definition 27: sequences, range and terms

- A **sequence** is a function $a(n)$ whose domain is $\mathbb{N}$.
- The **range** of a sequence is the set of all distinct values of $a(n)$.
- The **terms** of a sequence are the values $a(1)$, $a(2)$, ..., which are usually denoted with subscripts as a_1, a_2, ...

A sequence $a(n)$ is often denoted as $\{a_n\}$.

Notation: We use $\mathbb{N}$ to describe the set of natural numbers, that is, the integers 1, 2, 3, ...

Definition: factorial

The expression $3!$ refers to the number $3 \cdot 2 \cdot 1 = 6$. In general,

$$n! = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1 \tag{8.1.1}$$

where n is a natural number. We define $0! = 1$. While this does not immediately make sense, it makes many mathematical formulas work properly.

Example 8.1.1: Listing terms of a sequence

List the first four terms of the following sequences.

1. $\{a_n\} = \left\{\frac{3^n}{n!}\right\}$
2. $\{a_n\} = \{4 + (-1)^n\}$
3. $\{a_n\} = \left\{\frac{(-1)^{n(n+1)/2}}{n^2}\right\}$

Solution

1. $a_1 = \frac{3^1}{1!} = 3$; $\quad a_2 = \frac{3^2}{2!} = \frac{9}{2}$; $\quad a_3 = \frac{3^3}{3!} = \frac{9}{2}$; $\quad a_4 = \frac{3^4}{4!} = \frac{27}{8}$
   We can plot the terms of a sequence with a scatter plot. The "x"-axis is used for the values of n, and the values of the terms are plotted on the y-axis. To visualize this sequence, see Figure 8.1(a).
2. $a_1 = 4 + (-1)^1 = 3$; $\quad a_2 = 4 + (-1)^2 = 5$; $\quad a_3 = 4 + (-1)^3 = 3$; $\quad a_4 = 4 + (-1)^4 = 5$. Note that the range of this sequence is finite, consisting of only the values 3 and 5. This sequence is plotted in Figure 8.1(b).
3. $$a_1 = \frac{(-1)^{1(2)/2}}{1^2} = -1; \qquad a_2 = \frac{(-1)^{2(3)/2}}{2^2} = -\frac{1}{4} \tag{8.1.2}$$
   $$a_3 = \frac{(-1)^{3(4)/2}}{3^2} = \frac{1}{9} \qquad a_4 = \frac{(-1)^{4(5)/2}}{4^2} = \frac{1}{16}; \tag{8.1.3}$$
   $$a_5 = \frac{(-1)^{5(6)/2}}{5^2} = -\frac{1}{25} \tag{8.1.4}$$
   .
   We gave one extra term to begin to show the pattern of signs is "$-$, $-$, $+$, $+$, $-$, $-$, ..., due to the fact that the exponent of -1 is a special quadratic. This sequence is plotted in Figure 8.1(c).

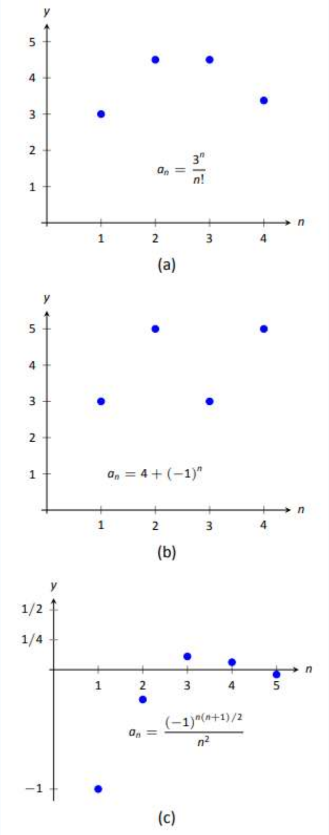


Figure 8.1: Plotting sequences in Example 8.1.1.

Example $8.1.2$: Determining a formula for a sequence

Find the n^{th} term of the following sequences, i.e., find a function that describes each of the given sequences.

1. 2, 5, 8, 11, 14, ...
2. 2, -5, 10, -17, 26, -37, ...
3. 1, 1, 2, 6, 24, 120, 720, ...
4. $\frac{5}{2}$, $\frac{5}{2}$, $\frac{15}{8}$, $\frac{5}{4}$, $\frac{25}{32}$, ...

Solution

We should first note that there is never exactly one function that describes a finite set of numbers as a sequence. There are many sequences that start with 2, then 5, as our first example does. We are looking for a simple formula that describes the terms given, knowing there is possibly more than one answer.

1. Note how each term is 3 more than the previous one. This implies a linear function would be appropriate: $a(n) = a_n = 3n + b$ for some appropriate value of b. As we want $a_1 = 2$, we set $b = -1$. Thus $a_n = 3n - 1$.
2. First notice how the sign changes from term to term. This is most commonly accomplished by multiplying the terms by either $(-1)^n$ or $(-1)^{n+1}$. Using $(-1)^n$ multiplies the odd terms by (-1); using $(-1)^{n+1}$ multiplies the even terms by (-1). As this sequence has negative even terms, we will multiply by $(-1)^{n+1}$.

   After this, we might feel a bit stuck as to how to proceed. At this point, we are just looking for a pattern of some sort: what do the numbers 2, 5, 10, 17, etc., have in common? There are many correct answers, but the one that we'll use here is that each is one more than a perfect square. That is, $2 = 1^1 + 1$, $5 = 2^2 + 1$, $10 = 3^2 + 1$, etc. Thus our formula is $a_n = (-1)^{n+1}(n^2 + 1)$.
3. One who is familiar with the factorial function will readily recognize these numbers. They are $0!$, $1!$, $2!$, $3!$, etc. Since our sequences start with $n = 1$, we cannot write $a_n = n!$, for this misses the $0!$ term. Instead, we shift by 1, and write $a_n = (n - 1)!$.

4. This one may appear difficult, especially as the first two terms are the same, but a little ``sleuthing" will help. Notice how the terms in the numerator are always multiples of 5, and the terms in the denominator are always powers of 2. Does something as simple as $a_n = \frac{5n}{2^n}$ work?

   When $n = 1$, we see that we indeed get $5/2$ as desired. When $n = 2$, we get $10/4 = 5/2$. Further checking shows that this formula indeed matches the other terms of the sequence.

A common mathematical endeavor is to create a new mathematical object (for instance, a sequence) and then apply previously known mathematics to the new object. We do so here. The fundamental concept of calculus is the limit, so we will investigate what it means to find the limit of a sequence.

Definition 28 LIMIT OF A SEQUENCE, CONVERGENT, DIVERGENT

Let $\{a_n\}$ be a sequence and let L be a real number. Given any $\epsilon > 0$, if an m can be found such that $|a_n - L| < \epsilon$ for all $n > m$, then we say the **limit** of $\{a_n\}$, **as** n **approaches infinity**, **is** L, denoted

$$\lim_{n\to\infty} a_n = L. \tag{8.1.5}$$

If $\lim_{n\to\infty} a_n$ exists, we say the sequence **converges**; otherwise, the sequence **diverges**.

This definition states, informally, that if the limit of a sequence is L, then if you go far enough out along the sequence, all subsequent terms will be *really close* to L. Of course, the terms "far enough" and "really close" are subjective terms, but hopefully the intent is clear.

This definition is reminiscent of the ϵ--δ proofs of Chapter 1. In that chapter we developed other tools to evaluate limits apart from the formal definition; we do so here as well.

tHEOREM 55: LIMIT OF A SEQUENCE

Let $\{a_n\}$ be a sequence and let $f(x)$ be a function whose domain contains the positive real numbers where $f(n) = a_n$ for all n in $\mathbb{N}$.

Theorem 55 allows us, in certain cases, to apply the tools developed in Chapter 1 to limits of sequences. Note two things *not* stated by the theorem:

1. If $\lim_{x\to\infty} f(x)$ does not exist, we cannot conclude that $\lim_{n\to\infty} a_n$ does not exist. It may, or may not, exist. For instance, we can define a sequence $\{a_n\} = \{\cos(2\pi n)\}$. Let $f(x) = \cos(2\pi x)$. Since the cosine function oscillates over the real numbers, the limit $\lim_{x\to\infty} f(x)$ does not exist.

   However, for every positive integer n, $\cos(2\pi n) = 1$, so $\lim_{n\to\infty} a_n = 1$.
2. If we cannot find a function $f(x)$ whose domain contains the positive real numbers where $f(n) = a_n$ for all n in $\mathbb{N}$, we cannot conclude $\lim_{n\to\infty} a_n$ does not exist. It may, or may not, exist.

Example $8.1.3$: Determining convergence/divergence of a sequence

Determine the convergence or divergence of the following sequences.

1. $\{a_n\} = \left\{\frac{3n^2-2n+1}{n^2-1000}\right\}$
2. $\{a_n\} = \{\cos n\}$
3. $\{a_n\} = \left\{\frac{(-1)^n}{n}\right\}$

Solution

1. Using Theorem 11, we can state that $\lim_{x\to\infty} \frac{3x^2-2x+1}{x^2-1000} = 3$. (We could have also directly applied l'H\^opital's Rule.) Thus the sequence $\{a_n\}$ converges, and its limit is 3. A scatter plot of every 5 values of a_n is given in Figure 8.2 (a). The values of a_n vary widely near $n = 30$, ranging from about -73 to 125, but as n grows, the values approach 3.

2. The limit $\lim_{x\to\infty} \cos x$ does not exist, as $\cos x$ oscillates (and takes on every value in $[-1,1]$ infinitely many times). Thus we cannot apply Theorem 55.

   The fact that the cosine function oscillates strongly hints that $\cos n$, when n is restricted to $\mathbb{N}$, will also oscillate. Figure 8.2 (b), where the sequence is plotted, shows that this is true. Because only discrete values of cosine are plotted, it does not bear strong resemblance to the familiar cosine wave.

   We conclude that $\lim_{n\to\infty} a_n$ does not exist.
3. We cannot actually apply Theorem 55 here, as the function $f(x)=(-1)^x/x$ is not well defined. (What does $(-1)^{\sqrt{2}}$ mean? In actuality, there is an answer, but it involves *complex analysis*, beyond the scope of this text.) So for now we say that we cannot determine the limit. (But we will be able to very soon.) By looking at the plot in Figure 8.2 (c), we would like to conclude that the sequence converges to 0. That is true, but at this point we are unable to decisively say so.

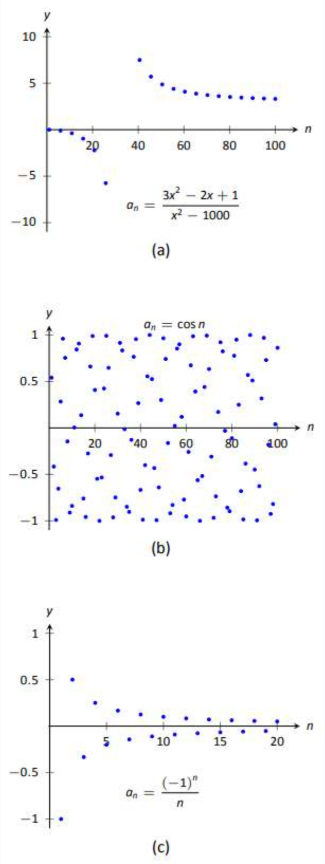


Figure 8.2: Scatter plots of the sequences in Example 8.1.2.

It seems that $\{(-1)^n/n\}$ converges to 0 but we lack the formal tool to prove it. The following theorem gives us that tool.

THEOREM 56: ABSOLUTE VALUE THEOREM

Let $\{a_n\}$ be a sequence. If $\lim_{n\to\infty} |a_n| = 0$, then $\lim_{n\to\infty} a_n = 0$

Example 8.1.4: Determining the convergence/divergence of a sequence

Determine the convergence or divergence of the following sequences.

1. $\{a_n\} = \left\{\frac{(-1)^n}{n}\right\}$
2. $\{a_n\} = \left\{\frac{(-1)^n(n+1)}{n}\right\}$

Solution

1. This appeared in Example 8.1.1. We want to apply Theorem 56, so consider the limit of $\{|a_n|\}$:

$$\begin{aligned}\lim_{n\to\infty}|a_n| &= \lim_{n\to\infty}\left|\frac{(-1)^n}{n}\right| \\ &= \lim_{n\to\infty}\frac{1}{n} \\ &= 0.\end{aligned}$$

   Since this limit is 0, we can apply Theorem 56 and state that $\lim\limits_{n\to\infty} a_n = 0$.

2. Because of the alternating nature of this sequence (i.e., every other term is multiplied by -1), we cannot simply look at the limit $\lim\limits_{x\to\infty}\frac{(-1)^x(x+1)}{x}$. We can try to apply the techniques of Theorem 56:

$$\begin{aligned}\lim_{n\to\infty}|a_n| &= \lim_{n\to\infty}\left|\frac{(-1)^n(n+1)}{n}\right| \\ &= \lim_{n\to\infty}\frac{n+1}{n} \\ &= 1.\end{aligned}$$

   We have concluded that when we ignore the alternating sign, the sequence approaches 1. This means we cannot apply Theorem 56; it states the the limit must be 0 in order to conclude anything.

   Since we know that the signs of the terms alternate *and* we know that the limit of $|a_n|$ is 1, we know that as n approaches infinity, the terms will alternate between values close to 1 and -1, meaning the sequence diverges. A plot of this sequence is given in Figure 8.3.

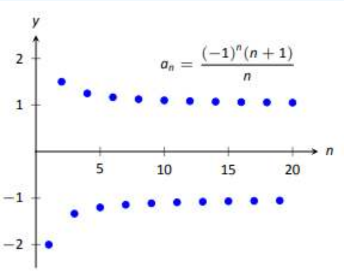


Figure 8.3: A plot of a sequence in Example 8.1.2, part 2.

We continue our study of the limits of sequences by considering some of the properties of these limits.

THEOREM 57: Properties of the Limits of Sequences

Let $\{a_n\}$ and $\{b_n\}$ be sequences such that $\lim\limits_{n\to\infty} a_n = L$, $s \lim\limits_{n\to\infty} b_n = K$, and let c be a real number.

1. $\lim\limits_{n\to\infty}(a_n \pm b_n) = L \pm K$
2. $\lim\limits_{n\to\infty}(a_n \cdot b_n) = L \cdot K$
3. $\lim\limits_{n\to\infty}(a_n/b_n) = L/K, K \neq 0$
4. $\lim\limits_{n\to\infty} c \cdot a_n = c \cdot L$

Example 8.1.5: Applying properties of limits of sequences

Let the following sequences, and their limits, be given:

- $\{a_n\} = \left\{\frac{n+1}{n^2}\right\}$, and $\lim\limits_{n\to\infty} a_n = 0$;
- $\{b_n\} = \left\{\left(1+\frac{1}{n}\right)^n\right\}$, and $\lim\limits_{n\to\infty} b_n = e$; and
- $\{c_n\} = \{n \cdot \sin(5/n)\}$, and $\lim\limits_{n\to\infty} c_n = 5$.

Evaluate the following limits.

1. $\lim\limits_{n\to\infty}(a_n + b_n)$
2. $\lim\limits_{n\to\infty}(b_n \cdot c_n)$
3. $\lim\limits_{n\to\infty}(1000 \cdot a_n)$

Solution

We will use Theorem 57 to answer each of these.

1. Since $\lim\limits_{n\to\infty} a_n = 0$ and $\lim\limits_{n\to\infty} b_n = e$, we conclude that $\lim\limits_{n\to\infty}(a_n + b_n) = 0 + e = e$. So even though we are adding something to each term of the sequence b_n, we are adding something so small that the final limit is the same as before.
2. Since $\lim\limits_{n\to\infty} b_n = e$ and $\lim\limits_{n\to\infty} c_n = 5$, we conclude that $\lim\limits_{n\to\infty}(b_n \cdot c_n) = e \cdot 5 = 5e$.
3. Since $\lim\limits_{n\to\infty} a_n = 0$, we have $\lim\limits_{n\to\infty} 1000a_n = 1000 \cdot 0 = 0$. It does not matter that we multiply each term by 1000; the sequence still approaches 0. (It just takes longer to get close to 0.)

There is more to learn about sequences than just their limits. We will also study their range and the relationships terms have with the terms that follow. We start with some definitions describing properties of the range.

Definition 29 Bounded and unbounded sequences

A sequence $\{a_n\}$ is said to be **bounded** if there exists real numbers m and M such that $m < a_n < M$ for all n in $\mathbb{N}$.

A sequence $\{a_n\}$ is said to be **unbounded** if it is not bounded.

A sequence $\{a_n\}$ is said to be **bounded above** if there exists an M such that $a_n < M$ for all n in $\mathbb{N}$; it is **bounded below** if there exists an m such that $m < a_n$ for all n in $\mathbb{N}$.

It follows from this definition that an unbounded sequence may be bounded above or bounded below; a sequence that is both bounded above and below is simply a bounded sequence.

Example 8.1.6: Determining boundedness of sequences

Determine the boundedness of the following sequences.

1. \(\{a_n\} = \left\{\frac1n\right\}\
2. $\{a_n\} = \{2^n\}$

Solution

1. The terms of this sequence are always positive but are decreasing, so we have $0 < a_n < 2$ for all n. Thus this sequence is bounded. Figure 8.4(a) illustrates this.
2. The terms of this sequence obviously grow without bound. However, it is also true that these terms are all positive, meaning $0 < a_n$. Thus we can say the sequence is unbounded, but also bounded below. Figure 8.4(b) illustrates this.

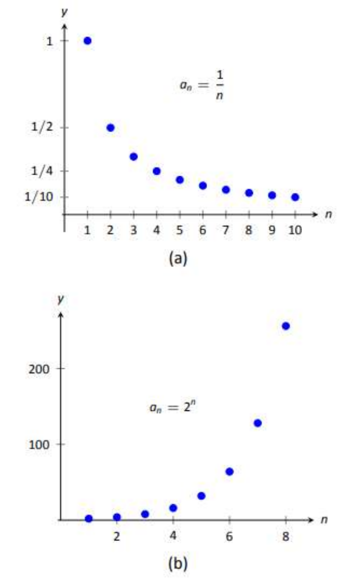


Figure 8.4: A plot of $a_n = 1/n$ and $a_n = 2^n$ from Example 8.1.6.

The previous example produces some interesting concepts. First, we can recognize that the sequence $\{1/n\}$ converges to 0. This says, informally, that "most" of the terms of the sequence are "really close" to 0. This implies that the sequence is bounded, using the following logic. First, "most" terms are near 0, so we could find some sort of bound on these terms (using Definition 28, the bound is ϵ). That leaves a "few" terms that are not near 0 (i.e., a *finite* number of terms). A finite list of numbers is always bounded.

This logic implies that if a sequence converges, it must be bounded. This is indeed true, as stated by the following theorem.

THEOREM 58 CONVERGENT SEQUENCES ARE BOUNDED

Let $\{a_n\}$ be a convergent sequence. Then $\{a_n\}$ is bounded.

In Example 8.1.5 we saw the sequence $\{b_n\} = \{(1+1/n)^n\}$, where it was stated that $\lim\limits_{n\to\infty} b_n = e$. (Note that this is simply restating part of Theorem 5.) Even though it may be difficult to intuitively grasp the behavior of this sequence, we know immediately that it is bounded.

Another interesting concept to come out of Example 8.1.6 again involves the sequence $\{1/n\}$. We stated, without proof, that the terms of the sequence were decreasing. That is, that $a_{n+1} < a_n$ for all n. (This is easy to show. Clearly $n < n+1$. Taking reciprocals flips the inequality: $1/n > 1/(n+1)$. This is the same as \(a_n > a_{n+1}$.) Sequences that either steadily increase or decrease are important, so we give this property a name.

Definition 30 MONOTONIC SEQUENCES

1. A sequence $\{a_n\}$ is **monotonically increasing** if $a_n \leq a_{n+1}$ for all n, i.e.,
$$a_1 \leq a_2 \leq a_3 \leq \cdots a_n \leq a_{n+1} \cdots \tag{8.1.6}$$
2. A sequence $\{a_n\}$ is **monotonically decreasing** if $a_n \geq a_{n+1}$ for all n, i.e.,
$$a_1 \geq a_2 \geq a_3 \geq \cdots a_n \geq a_{n+1} \cdots \tag{8.1.7}$$
3. A sequence is **monotonic** if it is monotonically increasing or monotonically decreasing.

NOTE: It is sometimes useful to call a monotonically increasing sequence *strictly increasing* if $a_n < a_{n+1}$ for all n; i.e, we remove the possibility that subsequent terms are equal. A similar statement holds for *strictly* decreasing.

Example 8.1.7: Determining monotonicity

Determine the monotonicity of the following sequences.

1. $\{a_n\} = \left\{\dfrac{n+1}{n}\right\}$
2. $\{a_n\} = \left\{\dfrac{n^2+1}{n+1}\right\}$
3. $\{a_n\} = \left\{\dfrac{n^2-9}{n^2-10n+26}\right\}$
4. $\{a_n\} = \left\{\dfrac{n^2}{n!}\right\}$

Solution

In each of the following, we will examine $a_{n+1} - a_n$. If $a_{n+1} - a_n > 0$, we conclude that $a_n < a_{n+1}$ and hence the sequence is increasing. If $a_{n+1} - a_n < 0$, we conclude that $a_n > a_{n+1}$ and the sequence is decreasing. Of course, a sequence need not be monotonic and perhaps neither of the above will apply.

We also give a scatter plot of each sequence. These are useful as they suggest a pattern of monotonicity, but analytic work should be done to confirm a graphical trend.

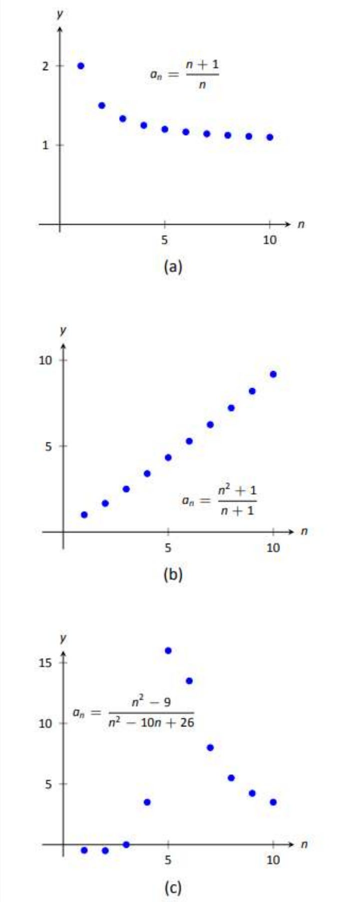


Figure 8.5: Plots of sequences in Example 8.1.7.

1. $$\begin{aligned} a_{n+1}-a_n &= \frac{n+2}{n+1}-\frac{n+1}{n} \quad (8.1.8)\\ &= \frac{(n+2)(n)-(n+1)^2}{(n+1)n} \quad (8.1.9)\\ &= \frac{-1}{n(n+1)} \quad (8.1.10)\\ &< 0 \quad \text{for all } n. \quad (8.1.11)\end{aligned}$$

   Since $a_{n+1}-a_n<0$ for all n, we conclude that the sequence is decreasing.

2. $$\begin{aligned} a_{n+1}-a_n &= \frac{(n+1)^2+1}{n+2}-\frac{n^2+1}{n+1} \quad (8.1.12)\\ &= \frac{\big((n+1)^2+1\big)(n+1)-(n^2+1)(n+2)}{(n+1)(n+2)} \quad (8.1.13)\\ &= \frac{n^2+4n+1}{(n+1)(n+2)} \quad (8.1.14)\\ &> 0 \quad \text{for all } n. \quad (8.1.15)\end{aligned}$$

   Since $a_{n+1}-a_n>0$ for all n, we conclude the sequence is increasing.

3. We can clearly see in Figure 8.5 (c), where the sequence is plotted, that it is not monotonic. However, it does seem that after the first 4 terms it is decreasing. To understand why, perform the same analysis as done before:

   $$\begin{aligned} a_{n+1}-a_n &= \frac{(n+1)^2-9}{(n+1)^2-10(n+1)+26}-\frac{n^2-9}{n^2-10n+26} \quad (8.1.16)\\ &= \frac{n^2+2n-8}{n^2-8n+17}-\frac{n^2-9}{n^2-10n+26} \quad (8.1.17)\end{aligned}$$

$$= \frac{(n^2+2n-8)(n^2-10n+26)-(n^2-9)(n^2-8n+17)}{(n^2-8n+17)(n^2-10n+26)} \tag{8.1.18}$$

$$= \frac{-10n^2+60n-55}{(n^2-8n+17)(n^2-10n+26)}. \tag{8.1.19}$$

We want to know when this is greater than, or less than, 0. The denominator is always positive, therefore we are only concerned with the numerator. Using the quadratic formula, we can determine that $-10n^2+60n-55=0$ when $n \approx 1.13, 4.87$. So for $n<1.13$, the sequence is decreasing. Since we are only dealing with the natural numbers, this means that $a_1 > a_2$.

Between 1.13 and 4.87, i.e., for $n=2$, 3 and 4, we have that $a_{n+1} > a_n$ and the sequence is increasing. (That is, when $n=2$, 3 and 4, the numerator $-10n^2+60n+55$ from the fraction above is >0.)

When $n>4.87$, i.e, for $n \geq 5$, we have that $-10n^2+60n+55<0$, hence $a_{n+1}-a_n<0$, so the sequence is decreasing.

In short, the sequence is simply not monotonic. However, it is useful to note that for $n \geq 5$, the sequence is monotonically decreasing.

4. Again, the plot in Figure 8.6 shows that the sequence is not monotonic, but it suggests that it is monotonically decreasing after the first term. We perform the usual analysis to confirm this.

$$a_{n+1}-a_n = \frac{(n+1)^2}{(n+1)!} - \frac{n^2}{n!} \tag{8.1.20}$$

$$= \frac{(n+1)^2 - n^2(n+1)}{(n+1)!} \tag{8.1.21}$$

$$= \frac{-n^3+2n+1}{(n+1)!} \tag{8.1.22}$$

When $n=1$, the above expression is >0; for $n \geq 2$, the above expression is <0. Thus this sequence is not monotonic, but it is monotonically decreasing after the first term.

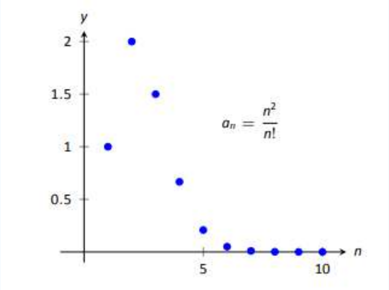


Figure 8.6: A plot of $a_n = n^2/n!$ in Example 8.1.7.

Knowing that a sequence is monotonic can be useful. In particular, if we know that a sequence is bounded and monotonic, we can conclude it converges! Consider, for example, a sequence that is monotonically decreasing and is bounded below. We know the sequence is always getting smaller, but that there is a bound to how small it can become. This is enough to prove that the sequence will converge, as stated in the following theorem.

THEOREM 59 BOUNDED MONOTONIC SEQUENCES ARE CONVERGENT

1. Let $\{a_n\}$ be a bounded, monotonic sequence. Then $\{a_n\}$ converges; i.e., $\lim\limits_{n\to\infty} a_n$ exists.
2. Let $\{a_n\}$ be a monotonically increasing sequence that is bounded above. Then $\{a_n\}$ converges.
3. Let $\{a_n\}$ be a monotonically decreasing sequence that is bounded below. Then $\{a_n\}$ converges.

Consider once again the sequence $\{a_n\} = \{1/n\}$. It is easy to show it is monotonically decreasing and that it is always positive (i.e., bounded below by 0). Therefore we can conclude by Theorem 59 that the sequence converges. We already knew this by other means, but in the following section this theorem will become very useful.

Sequences are a great source of mathematical inquiry. The On-Line Encyclopedia of Integer Sequences http://oeis.org contains thousands of sequences and their formulae. (As of this writing, there are 257,537 sequences in the database.) Perusing this database quickly demonstrates that a single sequence can represent several different ``real life" phenomena.

Interesting as this is, our interest actually lies elsewhere. We are more interested in the *sum* of a sequence. That is, given a sequence $\{a_n\}$, we are very interested in $a_1 + a_2 + a_3 + \cdots$. Of course, one might immediately counter with "Doesn't this just add up to `infinity'?" Many times, yes, but there are many important cases where the answer is no. This is the topic of *series*, which we begin to investigate in the next section.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 8.1: Sequences is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

8.2: Infinite Series

Given the sequence $\{a_n\} = \{1/2^n\} = 1/2,\ 1/4,\ 1/8,\ \ldots$, consider the following sums:

$$\begin{array}{ccccc} a_1 & = & 1/2 & = & 1/2 \\ a_1+a_2 & = & 1/2+1/4 & = & 3/4 \\ a_1+a_2+a_3 & = & 1/2+1/4+1/8 & = & 7/8 \\ a_1+a_2+a_3+a_4 & = & 1/2+1/4+1/8+1/16 & = & 15/16 \end{array} \tag{8.2.1}$$

In general, we can show that

$$a_1+a_2+a_3+\cdots+a_n = \frac{2^n-1}{2^n} = 1-\frac{1}{2^n}. \tag{8.2.2}$$

Let S_n be the sum of the first n terms of the sequence $\{1/2^n\}$. From the above, we see that $S_1 = 1/2$, $S_2 = 3/4$, etc. Our formula at the end shows that $S_n = 1-1/2^n$.

Now consider the following limit:

$$\lim_{n\to\infty} S_n = \lim_{n\to\infty}\left(1-1/2^n\right) = 1. \tag{8.2.3}$$

This limit can be interpreted as saying something amazing: *the sum of all* the terms of the sequence $\{1/2^n\}$ is 1.} This example illustrates some interesting concepts that we explore in this section. We begin this exploration with some definitions.

Definition 31: Infinite Series, n^{th} Partial Sums, Convergence, Divergence

Let $\{a_n\}$ be a sequence.

1. The sum $\sum_{n=1}^{\infty} a_n$ is an **infinite series** (or, simply **series**).
2. Let $S_n = \sum_{i=1}^{n} a_i$; the sequence $\{S_n\}$ is the sequence of n^{th} **partial sums** of $\{a_n\}$.
3. If the sequence $\{S_n\}$ converges to L, we say the series $\sum_{n=1}^{\infty} a_n$ **converges** to L, and we write $\sum_{n=1}^{\infty} a_n = L$.
4. If the sequence $\{S_n\}$ diverges, the series $\sum_{n=1}^{\infty} a_n$ **diverges**.

Using our new terminology, we can state that the series $\sum_{n=1}^{\infty} 1/2^n$ converges, and $\sum_{n=1}^{\infty} 1/2^n = 1$.

We will explore a variety of series in this section. We start with two series that diverge, showing how we might discern divergence.

Example $8.2.1$: Showing series diverge

1. Let $\{a_n\} = \{n^2\}$. Show $\sum_{n=1}^{\infty} a_n$ diverges.
2. Let $\{b_n\} = \{(-1)^{n+1}\}$. Show $\sum_{n=1}^{\infty} b_n$ diverges.

Solution

1. Consider S_n, the n^{th} partial sum.

$$\begin{aligned} S_n &= a_1+a_2+a_3+\cdots+a_n \\ &= 1^2+2^2+3^2\cdots+n^2. \end{aligned}$$

By Theorem 37, this is

$$= \frac{n(n+1)(2n+1)}{6}. \tag{8.2.4}$$

Since $\lim_{n\to\infty} S_n = \infty$, we conclude that the series $\sum_{n=1}^{\infty} n^2$ diverges. It is instructive to write $\sum_{n=1}^{\infty} n^2 = \infty$ for this tells us *how* the series diverges: it grows without bound.

A scatter plot of the sequences $\{a_n\}$ and $\{S_n\}$ is given in Figure 8.7(a). The terms of $\{a_n\}$ are growing, so the terms of the partial sums $\{S_n\}$ are growing even faster, illustrating that the series diverges.

2. The sequence $\{b_n\}$ starts with 1, -1, 1, -1, ... Consider some of the partial sums S_n of $\{b_n\}$:

$$\begin{aligned} S_1 &= 1 \\ S_2 &= 0 \\ S_3 &= 1 \\ S_4 &= 0 \end{aligned}$$

This pattern repeats; we find that $S_n = \begin{cases} 1 & n \text{ is odd} \\ 0 & n \text{ is even} \end{cases}$

As $\{S_n\}$ oscillates, repeating 1, 0, 1, 0, ..., we conclude that $\lim\limits_{n\to\infty} S_n$ does not exist, hence $\sum\limits_{n=1}^{\infty} (-1)^{n+1}$ diverges.

A scatter plot of the sequence $\{b_n\}$ and the partial sums $\{S_n\}$ is given in Figure 8.7(b). When n is odd, $b_n = S_n$ so the marks for b_n are drawn oversized to show they coincide.

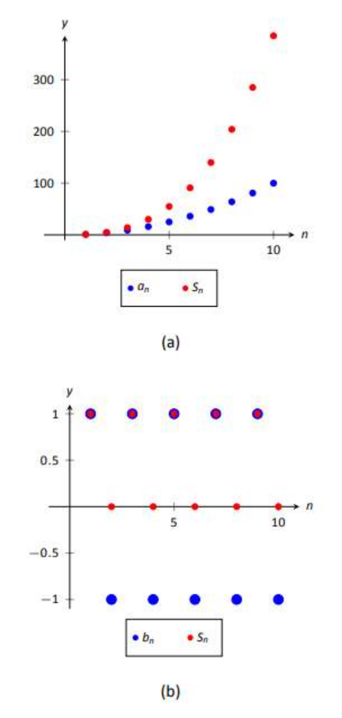


Figure 8.7: Scatter plots relating to Example 8.2.1.

While it is important to recognize when a series diverges, we are generally more interested in the series that converge. In this section we will demonstrate a few general techniques for determining convergence; later sections will delve deeper into this topic.

Geometric Series

One important type of series is a *geometric series*.

Definition 32: geometric series

A *geometric series* is a series of the form

$$\sum_{n=0}^{\infty} r^n = 1 + r + r^2 + r^3 + \cdots + r^n + \cdots \tag{8.2.5}$$

Note that the index starts at $n = 0$, not $n = 1$.

We started this section with a geometric series, although we dropped the first term of 1. One reason geometric series are important is that they have nice convergence properties.

theorem 60: convergence of geometric series

Consider the geometric series $\sum_{n=0}^{\infty} r^n$.

1. The n^{th} partial sum is: $S_n = \frac{1-r^{n+1}}{1-r}$.
2. The series converges if, and only if, $|r| < 1$. When $|r| < 1$,

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}. \tag{8.2.6}$$

According to Theorem 60, the series

$$\sum_{n=0}^{\infty} \frac{1}{2^n} = \sum_{n=0}^{\infty} \left(\frac{1}{2}\right)^2 = 1 + \frac{1}{2} + \frac{1}{4} + \cdots \tag{8.2.7}$$

converges as $r = 1/2$, and $\sum_{n=0}^{\infty} \frac{1}{2^n} = \frac{1}{1-1/2} = 2$. This concurs with our introductory example; while there we got a sum of 1, we skipped the first term of 1.

Example $8.2.2$: Exploring geometric series

Check the convergence of the following series. If the series converges, find its sum.

1. $\sum_{n=2}^{\infty} \left(\frac{3}{4}\right)^n$ 2. $\sum_{n=0}^{\infty} \left(\frac{-1}{2}\right)^n$ 3. $\sum_{n=0}^{\infty} 3^n$

Solution

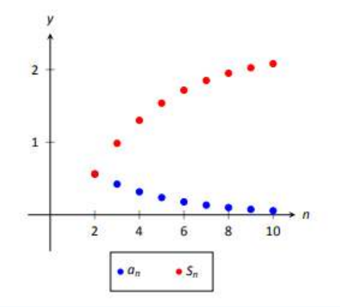


Figure 8.8: Scatter plots relating to the series in Example 8.2.2

1. Since $r = 3/4 < 1$, this series converges. By Theorem 60, we have that

$$\sum_{n=0}^{\infty} \left(\frac{3}{4}\right)^n = \frac{1}{1-3/4} = 4. \tag{8.2.8}$$

   However, note the subscript of the summation in the given series: we are to start with $n = 2$. Therefore we subtract off the first two terms, giving:

$$\sum_{n=2}^{\infty} \left(\frac{3}{4}\right)^n = 4 - 1 - \frac{3}{4} = \frac{9}{4}. \tag{8.2.9}$$

   This is illustrated in Figure 8.8.

2. Since $|r| = 1/2 < 1$, this series converges, and by Theorem 60,

$$\sum_{n=0}^{\infty} \left(\frac{-1}{2}\right)^n = \frac{1}{1-(-1/2)} = \frac{2}{3}. \tag{8.2.10}$$

The partial sums of this series are plotted in Figure 8.9(a). Note how the partial sums are not purely increasing as some of the terms of the sequence $\{(-1/2)^n\}$ are negative.

3. Since $r > 1$, the series diverges. (This makes "common sense"; we expect the sum

$$1+3+9+27+81+243+\cdots \tag{8.2.11}$$

to diverge.) This is illustrated in Figure 8.9(b).

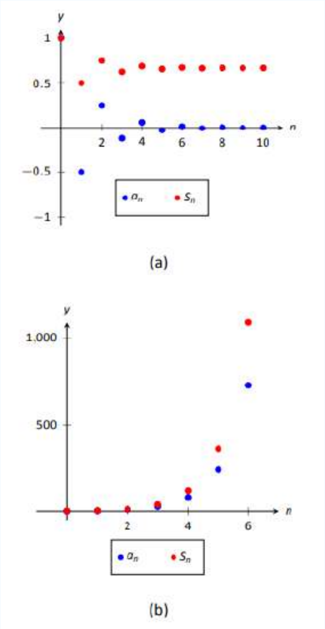


Figure 8.9: Scatter plots relating to the series in Example 8.2.2.

p-Series

Another important type of series is the *p-series*.

Definition 33: p-Series, General P-Series

1. A p--series is a series of the form

$$\sum_{n=1}^{\infty} \frac{1}{n^p}, \qquad \text{where } p > 0. \tag{8.2.12}$$

2. A **general** p--series} is a series of the form

$$\sum_{n=1}^{\infty} \frac{1}{(an+b)^p}, \qquad \text{where } p > 0 \text{ and } a,\, b \text{ are real numbers.} \tag{8.2.13}$$

Like geometric series, one of the nice things about p--series is that they have easy to determine convergence properties.

theorem 61: convergence of general P--Series

A general p--series $\sum\limits_{n=1}^{\infty} \frac{1}{(an+b)^p}$ will converge if, and only if, $p > 1$.

Note: Theorem 61 assumes that $an+b \neq 0$ for all n. If $an+b=0$ for some n, then of course the series does not converge regardless of p as not all of the terms of the sequence are defined.

Example 8.2.3: Determining convergence of series

Determine the convergence of the following series.

1. $\sum\limits_{n=1}^{\infty} \frac{1}{n}$
2. $\sum\limits_{n=1}^{\infty} \frac{1}{n^2}$
3. $\sum\limits_{n=1}^{\infty} \frac{1}{\sqrt{n}}$
4. $\sum\limits_{n=1}^{\infty} \frac{(-1)^n}{n}$
5. $\sum\limits_{n=11}^{\infty} \frac{1}{(\frac{1}{2}n-5)^3}$
6. $\sum\limits_{n=1}^{\infty} \frac{1}{2^n}$

Solution

1. This is a p--series with $p=1$. By Theorem 61, this series diverges.
   This series is a famous series, called the *Harmonic Series*, so named because of its relationship to *harmonics* in the study of music and sound.
2. This is a p--series with $p=2$. By Theorem 61, it converges. Note that the theorem does not give a formula by which we can determine *what* the series converges to; we just know it converges. A famous, unexpected result is that this series converges to $\pi^2/6$.
3. This is a p--series with $p=1/2$; the theorem states that it diverges.
4. This is not a p--series; the definition does not allow for alternating signs. Therefore we cannot apply Theorem 61. (Another famous result states that this series, the *Alternating Harmonic Series*, converges to $\ln 2$.)
5. This is a general p--series with $p=3$, therefore it converges.
6. This is not a p--series, but a geometric series with $r=1/2$. It converges.

Later sections will provide tests by which we can determine whether or not a given series converges. This, in general, is much easier than determining *what* a given series converges to. There are many cases, though, where the sum can be determined.

Example 8.2.4: Telescoping series

Evaluate the sum $\sum\limits_{n=1}^{\infty} \left(\frac{1}{n}-\frac{1}{n+1}\right)$.

Solution

It will help to write down some of the first few partial sums of this series.

$$\begin{aligned}
S_1 &= \frac{1}{1}-\frac{1}{2} & &= 1-\frac{1}{2}\\
S_2 &= \left(\frac{1}{1}-\frac{1}{2}\right)+\left(\frac{1}{2}-\frac{1}{3}\right) & &= 1-\frac{1}{3}\\
S_3 &= \left(\frac{1}{1}-\frac{1}{2}\right)+\left(\frac{1}{2}-\frac{1}{3}\right)+\left(\frac{1}{3}-\frac{1}{4}\right) & &= 1-\frac{1}{4}\\
S_4 &= \left(\frac{1}{1}-\frac{1}{2}\right)+\left(\frac{1}{2}-\frac{1}{3}\right)+\left(\frac{1}{3}-\frac{1}{4}\right)+\left(\frac{1}{4}-\frac{1}{5}\right) & &= 1-\frac{1}{5}
\end{aligned}$$

Note how most of the terms in each partial sum are canceled out! In general, we see that $S_n = 1-\frac{1}{n+1}$. The sequence $\{S_n\}$ converges, as $\lim\limits_{n\to\infty} S_n = \lim\limits_{n\to\infty}\left(1-\frac{1}{n+1}\right)=1$, and so we conclude that $\sum\limits_{n=1}^{\infty}\left(\frac{1}{n}-\frac{1}{n+1}\right)=1$. Partial sums of the series are plotted in Figure 8.10.

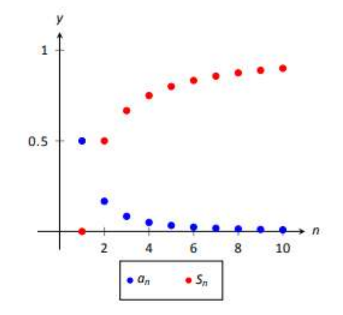


Figure 8.10: Scatter plots relating to the series of Example 8.2.4.

The series in Example 8.2.4 is an example of a **telescoping series**. Informally, a telescoping series is one in which the partial sums reduce to just a finite number of terms. The partial sum S_n did not contain n terms, but rather just two: 1 and $1/(n+1)$.

When possible, seek a way to write an explicit formula for the n^{th} partial sum S_n. This makes evaluating the limit $\lim\limits_{n\to\infty} S_n$ much more approachable. We do so in the next example.

Note on notation: Most of the series we encounter will start with $n=1$. For ease of notation, we will often write $\sum a_n$ instead of writing $\sum\limits_{n=1}^{\infty} a_n$.

Example 8.2.5: Evaluating series

Evaluate each of the following infinite series.

1. $\sum\limits_{n=1}^{\infty} \frac{2}{n^2+2n}$ 2. $\sum\limits_{n=1}^{\infty} \ln\left(\frac{n+1}{n}\right)$

Solution

1. We can decompose the fraction $2/(n^2+2n)$ as

$$\frac{2}{n^2+2n} = \frac{1}{n} - \frac{1}{n+2}. \tag{8.2.14}$$

(See Section 6.5, Partial Fraction Decomposition, to recall how this is done, if necessary.)
Expressing the terms of $\{S_n\}$ is now more instructive:

$$\begin{aligned}
S_1 &= 1-\frac{1}{3} & &= 1-\frac{1}{3}\\
S_2 &= \left(1-\frac{1}{3}\right)+\left(\frac{1}{2}-\frac{1}{4}\right) & &= 1+\frac{1}{2}-\frac{1}{3}-\frac{1}{4}\\
S_3 &= \left(1-\frac{1}{3}\right)+\left(\frac{1}{2}-\frac{1}{4}\right)+\left(\frac{1}{3}-\frac{1}{5}\right) & &= 1+\frac{1}{2}-\frac{1}{4}-\frac{1}{5}\\
S_4 &= \left(1-\frac{1}{3}\right)+\left(\frac{1}{2}-\frac{1}{4}\right)+\left(\frac{1}{3}-\frac{1}{5}\right)+\left(\frac{1}{4}-\frac{1}{6}\right) & &= 1+\frac{1}{2}-\frac{1}{5}-\frac{1}{6}\\
S_5 &= \left(1-\frac{1}{3}\right)+\left(\frac{1}{2}-\frac{1}{4}\right)+\left(\frac{1}{3}-\frac{1}{5}\right)+\left(\frac{1}{4}-\frac{1}{6}\right)+\left(\frac{1}{5}-\frac{1}{7}\right) & &= 1+\frac{1}{2}-\frac{1}{6}-\frac{1}{7}
\end{aligned}$$

We again have a telescoping series. In each partial sum, most of the terms cancel and we obtain the formula $S_n = 1+\frac{1}{2} - \frac{1}{n+1} - \frac{1}{n+2}$. Taking limits allows us to determine the convergence of the series:

$$\lim_{n\to\infty} S_n = \lim_{n\to\infty}\left(1+\frac{1}{2}-\frac{1}{n+1}-\frac{1}{n+2}\right) = \frac{3}{2}, \quad \text{so} \quad \sum_{n=1}^{\infty}\frac{1}{n^2+2n} = \frac{3}{2}. \tag{8.2.15}$$

This is illustrated in Figure 8.11(a).

2. We begin by writing the first few partial sums of the series:

$$\begin{aligned}S_1 &= \ln(2)\\ S_2 &= \ln(2)+\ln\left(\frac{3}{2}\right)\\ S_3 &= \ln(2)+\ln\left(\frac{3}{2}\right)+\ln\left(\frac{4}{3}\right)\\ S_4 &= \ln(2)+\ln\left(\frac{3}{2}\right)+\ln\left(\frac{4}{3}\right)+\ln\left(\frac{5}{4}\right)\end{aligned}$$

At first, this does not seem helpful, but recall the logarithmic identity: $\ln x + \ln y = \ln(xy)$. Applying this to S_4 gives:

$$S_4 = \ln(2)+\ln\left(\frac{3}{2}\right)+\ln\left(\frac{4}{3}\right)+\ln\left(\frac{5}{4}\right) = \ln\left(\frac{2}{1}\cdot\frac{3}{2}\cdot\frac{4}{3}\cdot\frac{5}{4}\right) = \ln(5). \tag{8.2.16}$$

We can conclude that $\{S_n\} = \{\ln(n+1)\}$. This sequence does not converge, as $\lim\limits_{n\to\infty} S_n = \infty$. Therefore $\sum\limits_{n=1}^{\infty} \ln\left(\frac{n+1}{n}\right) = \infty$; the series diverges. Note in Figure 8.11(b) how the sequence of partial sums grows slowly; after 100 terms, it is not yet over 5. Graphically we may be fooled into thinking the series converges, but our analysis above shows that it does not.

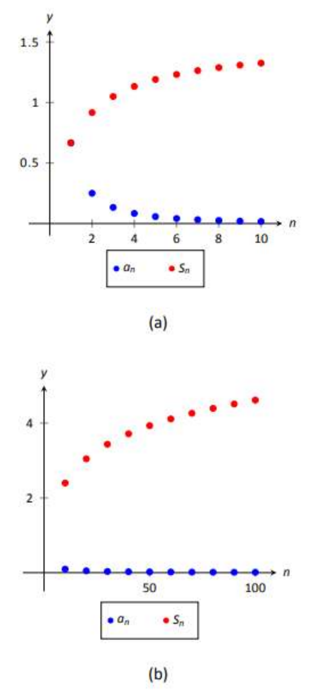


Figure 8.11: Scatter plots relating to the series in Example 8.2.5

We are learning about a new mathematical object, the series. As done before, we apply "old" mathematics to this new topic.

THEOREM 62 PROPERTIES OF INFINITE SERIES

Let $\sum\limits_{n=1}^{\infty} a_n = L$, $\sum\limits_{n=1}^{\infty} b_n = K$, and let c be a constant.

1. Constant Multiple Rule: $\sum\limits_{n=1}^{\infty} c\cdot a_n = c\cdot\sum\limits_{n=1}^{\infty} a_n = c\cdot L$.
2. Sum/Difference Rule: $\sum\limits_{n=1}^{\infty}\left(a_n \pm b_n\right) = \sum\limits_{n=1}^{\infty} a_n \pm \sum\limits_{n=1}^{\infty} b_n = L \pm K$.

Before using this theorem, we provide a few "famous" series.

KEY IDEA 31 IMPORTANT SERIES

1. $\sum_{n=0}^{\infty} \frac{1}{n!} = e$. (Note that the index starts with $n = 0$.)
2. $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$.
3. $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^2} = \frac{\pi^2}{12}$.
4. $\sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} = \frac{\pi}{4}$.
5. $\sum_{n=1}^{\infty} \frac{1}{n} \quad \text{diverges}$. (This is called the *Harmonic Series*.)
6. $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = \ln 2$. (This is called the *Alternating Harmonic Series*.)

Example $8.2.6$: Evaluating series

Evaluate the given series.

1. $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}(n^2-n)}{n^3}$ 2. $\sum_{n=1}^{\infty} \frac{1000}{n!}$ 3. $\frac{1}{16} + \frac{1}{25} + \frac{1}{36} + \frac{1}{49} + \cdots$

Solution

1. We start by using algebra to break the series apart:

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}(n^2-n)}{n^3} &= \sum_{n=1}^{\infty} \left(\frac{(-1)^{n+1}n^2}{n^3} - \frac{(-1)^{n+1}n}{n^3} \right) \\ &= \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} - \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^2} \\ &= \ln(2) - \frac{\pi^2}{12} \approx -0.1293. \end{aligned}$$

   This is illustrated in Figure 8.12(a).

2. This looks very similar to the series that involves e in Key Idea 31. Note, however, that the series given in this example starts with $n = 1$ and not $n = 0$. The first term of the series in the Key Idea is $1/0! = 1$, so we will subtract this from our result below:

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{1000}{n!} &= 1000 \cdot \sum_{n=1}^{\infty} \frac{1}{n!} \\ &= 1000 \cdot (e-1) \approx 1718.28. \end{aligned}$$

   This is illustrated in Figure 8.12(b). The graph shows how this particular series converges very rapidly.

3. The denominators in each term are perfect squares; we are adding $\sum_{n=4}^{\infty} \frac{1}{n^2}$ (note we start with $n = 4$, not $n = 1$). This series will converge. Using the formula from Key Idea 31, we have the following:

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{1}{n^2} &= \sum_{n=1}^{3} \frac{1}{n^2} + \sum_{n=4}^{\infty} \frac{1}{n^2} \\ \sum_{n=1}^{\infty} \frac{1}{n^2} - \sum_{n=1}^{3} \frac{1}{n^2} &= \sum_{n=4}^{\infty} \frac{1}{n^2} \\ \frac{\pi^2}{6} - \left(\frac{1}{1} + \frac{1}{4} + \frac{1}{9} \right) &= \sum_{n=4}^{\infty} \frac{1}{n^2} \\ \frac{\pi^2}{6} - \frac{49}{36} &= \sum_{n=4}^{\infty} \frac{1}{n^2} \\ 0.2838 &\approx \sum_{n=4}^{\infty} \frac{1}{n^2} \end{aligned}$$

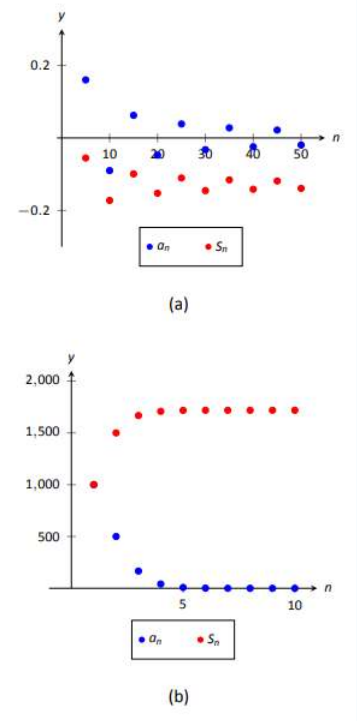


Figure 8.12: Scatter plots relating to the series in Example 8.2.6

It may take a while before one is comfortable with this statement, whose truth lies at the heart of the study of infinite series: *it is possible that the sum of an infinite list of nonzero numbers is finite*. We have seen this repeatedly in this section, yet it still may "take some getting used to."

As one contemplates the behavior of series, a few facts become clear.

1. In order to add an infinite list of nonzero numbers and get a finite result, "most" of those numbers must be "very near" 0.
2. If a series diverges, it means that the sum of an infinite list of numbers is not finite (it may approach $\pm\infty$ or it may oscillate), and:
   1. The series will still diverge if the first term is removed.
   2. The series will still diverge if the first 10 terms are removed.
   3. The series will still diverge if the first $1,000,000$ terms are removed.
   4. The series will still diverge if any finite number of terms from anywhere in the series are removed.

These concepts are very important and lie at the heart of the next two theorems.

theorem 63 n^{th}--Term Test for Convergence/Divergence

Consider the series $\sum\limits_{n=1}^{\infty} a_n$.

1. If $\sum\limits_{n=1}^{\infty} a_n$ converges, then $\lim\limits_{n\to\infty} a_n = 0$.
2. If $\lim\limits_{n\to\infty} a_n \neq 0$, then $\sum\limits_{n=1}^{\infty} a_n$ diverges.

Note that the two statements in Theorem 63 are really the same. In order to converge, the limit of the terms of the sequence must approach 0; if they do not, the series will not converge.

Looking back, we can apply this theorem to the series in Example 8.2.1. In that example, the n^{th} terms of both sequences do not converge to 0, therefore we can quickly conclude that each series diverges.

Important! This theorem *does not state* that if $\lim\limits_{n\to\infty} a_n = 0$ then $\sum\limits_{n=1}^{\infty} a_n$ converges. The standard example of this is the Harmonic Series, as given in Key Idea 31. The Harmonic Sequence, $\{1/n\}$, converges to 0; the Harmonic Series, $\sum\limits_{n=1}^{\infty} 1/n$, diverges.

theorem 64 infinite nature of series

The convergence or divergence remains unchanged by the addition or subtraction of any finite number of terms. That is:

1. A divergent series will remain divergent with the addition or subtraction of any finite number of terms.
2. A convergent series will remain convergent with the addition or subtraction of any finite number of terms. (Of course, the *sum* will likely change.)

Consider once more the Harmonic Series $\sum\limits_{n=1}^{\infty} \frac{1}{n}$ which diverges; that is, the sequence of partial sums $\{S_n\}$ grows (very, very slowly) without bound. One might think that by removing the "large" terms of the sequence that perhaps the series will converge. This is simply not the case. For instance, the sum of the first 10 million terms of the Harmonic Series is about 16.7. Removing the first 10 million terms from the Harmonic Series changes the n^{th} partial sums, effectively subtracting 16.7 from the sum. However, a sequence that is growing without bound will still grow without bound when 16.7 is subtracted from it.

The equations below illustrate this. The first line shows the infinite sum of the Harmonic Series split into the sum of the first 10 million terms plus the sum of "everything else." The next equation shows us subtracting these first 10 million terms from both sides. The final equation employs a bit of "psuedo--math": subtracting 16.7 from "infinity" still leaves one with "infinity."

$$\begin{aligned} \sum_{n=1}^{\infty}\frac{1}{n} &= \sum_{n=1}^{10{,}000{,}000}\frac{1}{n} \quad + \sum_{n=10{,}000{,}001}^{\infty}\frac{1}{n} \\ \sum_{n=1}^{\infty}\frac{1}{n} - \sum_{n=1}^{10{,}000{,}000}\frac{1}{n} &= \sum_{n=10{,}000{,}001}^{\infty}\frac{1}{n} \\ \infty - 16.7 &= \infty. \end{aligned}$$

This section introduced us to series and defined a few special types of series whose convergence properties are well known: we know when a p-series or a geometric series converges or diverges. Most series that we encounter are not one of these types, but we are still interested in knowing whether or not they converge. The next three sections introduce tests that help us determine whether or not a given series converges.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 8.2: Infinite Series is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

8.3: Integral and Comparison Tests

Knowing whether or not a series converges is very important, especially when we discusses Power Series. Theorems 60 and 61 give criteria for when Geometric and p-series converge, and Theorem 63 gives a quick test to determine if a series diverges. There are many important series whose convergence cannot be determined by these theorems, though, so we introduce a set of tests that allow us to handle a broad range of series. We start with the Integral Test.

Integral Test

We stated in Section 8.1 that a sequence $\{a_n\}$ is a function $a(n)$ whose domain is $\mathbb{N}$, the set of natural numbers. If we can extend $a(n)$ to $\mathbb{R}$, the real numbers, and it is both positive and decreasing on $[1,\infty)$, then the convergence of $\sum\limits_{n=1}^{\infty} a_n$ is the same as $\int\limits_1^{\infty} a(x)dx$.

theorem 8.3.1: integral test

Let a sequence $\{a_n\}$ be defined by $a_n = a(n)$, where $a(n)$ is continuous, positive and decreasing on $[1,\infty)$. Then $\sum\limits_{n=1}^{\infty} a_n$ converges, if, and only if, $\int\limits_1^{\infty} a(x)dx$ converges.

We can demonstrate the truth of the Integral Test with two simple graphs. In Figure $8.3.1a$, the height of each rectangle is $a(n) = a_n$ for $n = 1, 2, \ldots$, and clearly the rectangles enclose more area than the area under $y = a(x)$. Therefore we can conclude that

$$\int_1^{\infty} a(x)dx < \sum_{n=1}^{\infty} a_n. \tag{8.3.1}$$

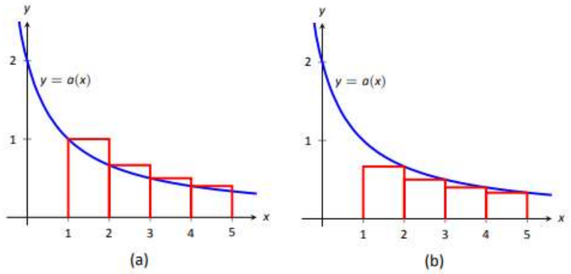


Figure $8.3.1$: Illustrating the truth of the Integral Test.

In Figure $8.3.1b$, we draw rectangles under $y = a(x)$ with the Right-Hand rule, starting with $n = 2$. This time, the area of the rectangles is less than the area under $y = a(x)$, so $\sum\limits_{n=2}^{\infty} a_n < \int\limits_1^{\infty} a(x)dx$. Note how this summation starts with $n = 2$; adding a_1 to both sides lets us rewrite the summation starting with $n = 1$:

$$\sum_{n=1}^{\infty} a_n < a_1 + \int_1^{\infty} a(x)dx. \tag{8.3.2}$$

Combining Equations 8.3.1 and 8.3.2, we have

$$\sum_{n=1}^{\infty} a_n < a_1 + \int_1^{\infty} a(x)dx < a_1 + \sum_{n=1}^{\infty} a_n. \tag{8.3.3}$$

Theorem $8.3.1$

From Equation 8.3.3 we can make the following two statements:

1. If $\sum\limits_{n=1}^{\infty} a_n$ diverges, so does $\int\limits_1^{\infty} a(x)dx$ (because $\sum\limits_{n=1}^{\infty} a_n < a_1 + \int\limits_1^{\infty} a(x)dx$)
2. If $\sum\limits_{n=1}^{\infty} a_n$ converges, so does $\int\limits_1^{\infty} a(x)dx$ (because $\int\limits_1^{\infty} a(x)dx < \sum\limits_{n=1}^{\infty} a_n$.)

Therefore the series and integral either both converge or both diverge.

Theorem$8.3.1$ allows us to extend this theorem to series where $a(n)$ is positive and decreasing on $[b,\infty)$ for some $b>1$.

Example $8.3.1$: Using the Integral Test

Determine the convergence of $\sum\limits_{n=1}^{\infty}\frac{\ln n}{n^2}$. (The terms of the sequence $\{a_n\}=\{\ln n/n^2\}$ and the n$^{\text{th}}$ partial sums are given in Figure $8.3.2$).

Solution

Figure $8.3.2$ implies that $a(n)=(\ln n)/n^2$ is positive and decreasing on $[2,\infty)$. We can determine this analytically, too. We know $a(n)$ is positive as both $\ln n$ and n^2 are positive on $[2,\infty)$. To determine that $a(n)$ is decreasing, consider $a'(n)=(1-2\ln n)/n^3$, which is negative for $n\geq 2$. Since $a'(n)$ is negative, $a(n)$ is decreasing.

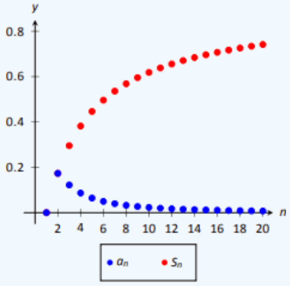


Figure $8.3.2$: Plotting the sequence and series in Example $8.3.1$.

Applying the Integral Test, we test the convergence of $\int\limits_1^{\infty}\frac{\ln x}{x^2}dx$. Integrating this improper integral requires the use of Integration by Parts, with $u=\ln x$ and $dv=1/x^2dx$.

$$\begin{aligned}\int\limits_1^{\infty}\frac{\ln x}{x^2}dx &= \lim_{b\to\infty}\int\limits_1^{b}\frac{\ln x}{x^2}dx\\ &= \lim_{b\to\infty} -\frac{1}{x}\ln x\Big|_1^b+\int\limits_1^b\frac{1}{x^2}dx\\ &= \lim_{b\to\infty} -\frac{1}{x}\ln x-\frac{1}{x}\Big|_1^b\\ &= \lim_{b\to\infty} 1-\frac{1}{b}-\frac{\ln b}{b}. \quad \text{Apply L'Hôpital's Rule:}\\ &= 1.\end{aligned}$$

Since $\int\limits_1^{\infty}\frac{\ln x}{x^2}dx$ converges, so does $\sum\limits_{n=1}^{\infty}\frac{\ln n}{n^2}$.

Theorem 61 was given without justification, stating that the general p-series $\sum\limits_{n=1}^{\infty}\frac{1}{(an+b)^p}$ converges if, and only if, $p>1$. In the following example, we prove this to be true by applying the Integral Test.

Example $8.3.2$: Using the Integral Test to establish Theorem 61

Use the Integral Test to prove that $\sum\limits_{n=1}^{\infty}\frac{1}{(an+b)^p}$ converges if, and only if, $p>1$.

Solution

Consider the integral $\int\limits_1^{\infty}\frac{1}{(ax+b)^p}dx$; assuming $p\neq 1$,

$$\begin{aligned}\int_1^\infty \frac{1}{(ax+b)^p}\,dx &= \lim_{c\to\infty}\int_1^c \frac{1}{(ax+b)^p}\,dx\\ &= \lim_{c\to\infty}\frac{1}{a(1-p)}(ax+b)^{1-p}\Big|_1^c\\ &= \lim_{c\to\infty}\frac{1}{a(1-p)}\left((ac+b)^{1-p}-(a+b)^{1-p}\right).\end{aligned}$$

This limit converges if, and only if, $p>1$. It is easy to show that the integral also diverges in the case of $p=1$. (This result is similar to the work preceding Key Idea 21.)

Therefore $\sum_{n=1}^{\infty}\frac{1}{(an+b)^p}$ converges if, and only if, $p>1$.

We consider two more convergence tests in this section, both *comparison* tests. That is, we determine the convergence of one series by comparing it to another series with known convergence.

Direct Comparison Test

theorem $8.3.1$: direct comparison test

Let $\{a_n\}$ and $\{b_n\}$ be positive sequences where $a_n \le b_n$ for all $n\ge N$, for some $N\ge 1$.

1. If $\sum_{n=1}^{\infty} b_n$ converges, then $\sum_{n=1}^{\infty} a_n$ converges.
2. If $\sum_{n=1}^{\infty} a_n$ diverges, then $\sum_{n=1}^{\infty} b_n$ diverges.

Note: A sequence $\{a_n\}$ is a **positive sequence** if $a_n>0$ for all n.

Because of Theorem 64, any theorem that relies on a positive sequence still holds true when $a_n>0$ for all but a finite number of values of n.

Example $8.3.3$: Applying the Direct Comparison Test

Determine the convergence of $\sum_{n=1}^{\infty}\frac{1}{3^n+n^2}$.

Solution

This series is neither a geometric or p-series, but seems related. We predict it will converge, so we look for a series with larger terms that converges. (Note too that the Integral Test seems difficult to apply here.)

Since $3^n<3^n+n^2$, $\frac{1}{3^n}>\frac{1}{3^n+n^2}$ for all $n\ge 1$. The series $\sum_{n=1}^{\infty}\frac{1}{3^n}$ is a convergent geometric series; by Theorem 66, $\sum_{n=1}^{\infty}\frac{1}{3^n+n^2}$ converges.

Example $8.3.4$: Applying the Direct Comparison Test

Determine the convergence of $\sum_{n=1}^{\infty}\frac{1}{n-\ln n}$.

Solution

We know the Harmonic Series $\sum_{n=1}^{\infty}\frac{1}{n}$ diverges, and it seems that the given series is closely related to it, hence we predict it will diverge.

Since $n\ge n-\ln n$ for all $n\ge 1$, $\frac{1}{n}\le\frac{1}{n-\ln n}$ for all $n\ge 1$.

The Harmonic Series diverges, so we conclude that $\sum_{n=1}^{\infty}\frac{1}{n-\ln n}$ diverges as well.

The concept of direct comparison is powerful and often relatively easy to apply. Practice helps one develop the necessary intuition to quickly pick a proper series with which to compare. However, it is easy to construct a series for which it is difficult to apply the

Direct Comparison Test.

Consider $\sum_{n=1}^{\infty} \frac{1}{n+\ln n}$. It is very similar to the divergent series given in Example 8.3.5. We suspect that it also diverges, as $\frac{1}{n} \approx \frac{1}{n+\ln n}$ for large n. However, the inequality that we naturally want to use "goes the wrong way": since $n \leq n+\ln n$ for all $n \geq 1$, $\frac{1}{n} \geq \frac{1}{n+\ln n}$ for all $n \geq 1$. The given series has terms *less than* the terms of a divergent series, and we cannot conclude anything from this.

Fortunately, we can apply another test to the given series to determine its convergence.

Large Limit Comparison Test

Theorem 67: limit comparison test

Let $\{a_n\}$ and $\{b_n\}$ be positive sequences.

1. If $\lim_{n\to\infty} \frac{a_n}{b_n} = L$, where L is a positive real number, then $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ either both converge or both diverge.
2. If $\lim_{n\to\infty} \frac{a_n}{b_n} = 0$, then if $\sum_{n=1}^{\infty} b_n$ converges, then so does $\sum_{n=1}^{\infty} a_n$.
3. If $\lim_{n\to\infty} \frac{a_n}{b_n} = \infty$, then if $\sum_{n=1}^{\infty} b_n$ diverges, then so does $\sum_{n=1}^{\infty} a_n$.

Theorem 67 is most useful when the convergence of the series from $\{b_n\}$ is known and we are trying to determine the convergence of the series from $\{a_n\}$.

We use the Limit Comparison Test in the next example to examine the series $\sum_{n=1}^{\infty} \frac{1}{n+\ln n}$ which motivated this new test.

Example $8.3.5$: Applying the Limit Comparison Test

Determine the convergence of $\sum_{n=1}^{\infty} \frac{1}{n+\ln n}$ using the Limit Comparison Test.

Solution

We compare the terms of $\sum_{n=1}^{\infty} \frac{1}{n+\ln n}$ to the terms of the Harmonic Sequence $\sum_{n=1}^{\infty} \frac{1}{n}$:

$$\begin{aligned} \lim_{n\to\infty} \frac{1/(n+\ln n)}{1/n} &= \lim_{n\to\infty} \frac{n}{n+\ln n} \\ &= 1 \quad \text{(after applying L'Hôpital's Rule)}. \end{aligned}$$

Since the Harmonic Series diverges, we conclude that $\sum_{n=1}^{\infty} \frac{1}{n+\ln n}$ diverges as well.

Example $8.3.6$: Applying the Limit Comparison Test

Determine the convergence of $\sum_{n=1}^{\infty} \frac{1}{3^n-n^2}$

Solution

This series is similar to the one in Example 8.3.3, but now we are considering "3^n-n^2" instead of "3^n+n^2." This difference makes applying the Direct Comparison Test difficult.

Instead, we use the Limit Comparison Test and compare with the series $\sum_{n=1}^{\infty} \frac{1}{3^n}$:

$$\begin{aligned} \lim_{n\to\infty} \frac{1/(3^n-n^2)}{1/3^n} &= \lim_{n\to\infty} \frac{3^n}{3^n-n^2} \\ &= 1 \quad \text{(after applying L'Hôpital's Rule twice)}. \end{aligned}$$

We know $\sum_{n=1}^{\infty} \frac{1}{3^n}$ is a convergent geometric series, hence $\sum_{n=1}^{\infty} \frac{1}{3^n-n^2}$ converges as well.

As mentioned before, practice helps one develop the intuition to quickly choose a series with which to compare. A general rule of thumb is to pick a series based on the dominant term in the expression of $\{a_n\}$. It is also helpful to note that factorials dominate exponentials, which dominate algebraic functions (e.g., polynomials), which dominate logarithms. In the previous example, the dominant term of $\frac{1}{3^n - n^2}$ was 3^n, so we compared the series to $\sum\limits_{n=1}^{\infty} \frac{1}{3^n}$. It is hard to apply the Limit Comparison Test to series containing factorials, though, as we have not learned how to apply L'H$\hat{o}$pital's Rule to $n!$.

Example $8.3.7$: Applying the Limit Comparison Test

Determine the convergence of $\sum\limits_{n=1}^{\infty} \frac{\sqrt{n}+3}{n^2-n+1}$.

Solution

We naively attempt to apply the rule of thumb given above and note that the dominant term in the expression of the series is $1/n^2$. Knowing that $\sum\limits_{n=1}^{\infty} \frac{1}{n^2}$ converges, we attempt to apply the Limit Comparison Test:

$$\begin{aligned} \lim_{n\to\infty} \frac{(\sqrt{n}+3)/(n^2-n+1)}{1/n^2} &= \lim_{n\to\infty} \frac{n^2(\sqrt{n}+3)}{n^2-n+1} \\ &= \infty \quad (\text{Apply L'Hôpital's Rule}). \end{aligned}$$

Theorem 67 part (3) only applies when $\sum\limits_{n=1}^{\infty} b_n$ diverges; in our case, it converges. Ultimately, our test has not revealed anything about the convergence of our series.

The problem is that we chose a poor series with which to compare. Since the numerator and denominator of the terms of the series are both algebraic functions, we should have compared our series to the dominant term of the numerator divided by the dominant term of the denominator.

The dominant term of the numerator is $n^{1/2}$ and the dominant term of the denominator is n^2. Thus we should compare the terms of the given series to $n^{1/2}/n^2 = 1/n^{3/2}$:

$$\begin{aligned} \lim_{n\to\infty} \frac{(\sqrt{n}+3)/(n^2-n+1)}{1/n^{3/2}} &= \lim_{n\to\infty} \frac{n^{3/2}(\sqrt{n}+3)}{n^2-n+1} \\ &= 1 \quad (\text{Apply L'Hôpital's Rule}). \end{aligned}$$

Since the p-series $\sum\limits_{n=1}^{\infty} \frac{1}{n^{3/2}}$ converges, we conclude that $\sum\limits_{n=1}^{\infty} \frac{\sqrt{n}+3}{n^2-n+1}$ converges as well.

We mentioned earlier that the Integral Test did not work well with series containing factorial terms. The next section introduces the **Ratio Test**, which does handle such series well. We also introduce the **Root Test**, which is good for series where each term is raised to a power.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 8.3: Integral and Comparison Tests is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

8.4: Ratio and Root Tests

The n^{th}--Term Test of Theorem 63 states that in order for a series $\sum\limits_{n=1}^{\infty} a_n$ to converge, $\lim\limits_{n\to\infty} a_n = 0$. That is, the terms of $\{a_n\}$ must get very small. Not only must the terms approach 0, they must approach 0 "fast enough": while $\lim\limits_{n\to\infty} 1/n = 0$, the Harmonic Series $\sum\limits_{n=1}^{\infty} \frac{1}{n}$ diverges as the terms of $\{1/n\}$ do not approach 0 "fast enough."

The comparison tests of the previous section determine convergence by comparing terms of a series to terms of another series whose convergence is known. This section introduces the Ratio and Root Tests, which determine convergence by analyzing the terms of a series to see if they approach 0 "fast enough."

Ratio Test

theorem 68: ratio test

Let $\{a_n\}$ be a positive sequence where $\lim\limits_{n\to\infty} \frac{a_{n+1}}{a_n} = L$.

1. If $L < 1$, then $\sum\limits_{n=1}^{\infty} a_n$ converges.
2. If $L > 1$ or $L = \infty$, then $\sum\limits_{n=1}^{\infty} a_n$ diverges.
3. If $L = 1$, the Ratio Test is inconclusive.

Theorem 64 allows us to apply the Ratio Test to series where $\{a_n\}$ is positive for all but a finite number of terms.

The principle of the Ratio Test is this: if $\lim\limits_{n\to\infty} \frac{a_{n+1}}{a_n} = L < 1$, then for large n, each term of $\{a_n\}$ is significantly smaller than its previous term which is enough to ensure convergence.

Example $8.4.1$: Applying the Ratio Test

Use the Ratio Test to determine the convergence of the following series:

1. $\sum\limits_{n=1}^{\infty} \frac{2^n}{n!}$.
2. $\sum\limits_{n=1}^{\infty} \frac{3^n}{n^3}$
3. $\sum\limits_{n=1}^{\infty} \frac{1}{n^2+1}$.

Solution

1. $\sum\limits_{n=1}^{\infty} \frac{2^n}{n!}$:

$$\begin{aligned} \lim_{n\to\infty} \frac{2^{n+1}/(n+1)!}{2^n/n!} &= \lim_{n\to\infty} \frac{2^{n+1}n!}{2^n(n+1)!} \\ &= \lim_{n\to\infty} \frac{2}{n+1} \\ &= 0. \end{aligned}$$

Since the limit is $0 < 1$, by the Ratio Test $\sum\limits_{n=1}^{\infty} \frac{2^n}{n!}$ converges.

2. $\sum\limits_{n=1}^{\infty} \frac{3^n}{n^3}$:

$$\begin{aligned} \lim_{n\to\infty} \frac{3^{n+1}/(n+1)^3}{3^n/n^3} &= \lim_{n\to\infty} \frac{3^{n+1}n^3}{3^n(n+1)^3} \\ &= \lim_{n\to\infty} \frac{3n^3}{(n+1)^3} \\ &= 3. \end{aligned}$$

Since the limit is $3 > 1$, by the Ratio Test $\sum\limits_{n=1}^{\infty} \frac{3^n}{n^3}$ diverges.

3. $\sum\limits_{n=1}^{\infty} \frac{1}{n^2+1}$:

$$\begin{aligned}\lim_{n\to\infty} \frac{1/\big((n+1)^2+1\big)}{1/(n^2+1)} &= \lim_{n\to\infty} \frac{n^2+1}{(n+1)^2+1} \\ &= 1.\end{aligned}$$

Since the limit is 1, the Ratio Test is inconclusive. We can easily show this series converges using the Direct or Limit Comparison Tests, with each comparing to the series $\sum\limits_{n=1}^{\infty} \frac{1}{n^2}$.

The Ratio Test is not effective when the terms of a series *only* contain algebraic functions (e.g., polynomials). It is most effective when the terms contain some factorials or exponentials. The previous example also reinforces our developing intuition: factorials dominate exponentials, which dominate algebraic functions, which dominate logarithmic functions. In Part 1 of the example, the factorial in the denominator dominated the exponential in the numerator, causing the series to converge. In Part 2, the exponential in the numerator dominated the algebraic function in the denominator, causing the series to diverge.

While we have used factorials in previous sections, we have not explored them closely and one is likely to not yet have a strong intuitive sense for how they behave. The following example gives more practice with factorials.

Example $8.4.2$: Applying the Ratio Test

Determine the convergence of $\sum\limits_{n=1}^{\infty} \frac{n!n!}{(2n)!}$.

Solution

Before we begin, be sure to note the difference between $(2n)!$ and $2n!$. When $n=4$, the former is $8! = 8\cdot 7\cdot \ldots \cdot 2\cdot 1 = 40,320$, whereas the latter is $2(4\cdot 3\cdot 2\cdot 1) = 48$.

Applying the Ratio Test:

$$\begin{aligned}\lim_{n\to\infty} \frac{(n+1)!(n+1)!/\big(2(n+1)\big)!}{n!n!/(2n)!} &= \lim_{n\to\infty} \frac{(n+1)!(n+1)!(2n)!}{n!n!(2n+2)!} \\ \text{Noting that } (2n+2)! = (2n+2)\cdot(2n+1)\cdot(2n)!\text{, we have} & \\ &= \lim_{n\to\infty} \frac{(n+1)(n+1)}{(2n+2)(2n+1)} \\ &= 1/4.\end{aligned}$$

Since the limit is $1/4 < 1$, by the Ratio Test we conclude $\sum\limits_{n=1}^{\infty} \frac{n!n!}{(2n)!}$ converges.

Root Test

The final test we introduce is the Root Test, which works particularly well on series where each term is raised to a power, and does not work well with terms containing factorials.

theorem 69: root test

Let $\{a_n\}$ be a positive sequence and let $\lim\limits_{n\to\infty} (a_n)^{1/n} = L$.

1. If $L < 1$, then $\sum\limits_{n=1}^{\infty} a_n$ converges.
2. If $L > 1$ or $L = \infty$, then $\sum\limits_{n=1}^{\infty} a_n$ diverges.
3. If $L = 1$, the Root Test is inconclusive.

Example $8.4.3$: Applying the Root Test

Determine the convergence of the following series using the Root Test:

1. $\sum\limits_{n=1}^{\infty} \left(\frac{3n+1}{5n-2}\right)^n$
2. $\sum\limits_{n=1}^{\infty} \frac{n^4}{(\ln n)^n}$
3. $\sum\limits_{n=1}^{\infty} \frac{2^n}{n^2}$.

Solution

1. $\lim_{n\to\infty}\left(\left(\frac{3n+1}{5n-2}\right)^n\right)^{1/n}=\lim_{n\to\infty}\frac{3n+1}{5n-2}=\frac{3}{5}.$
   Since the limit is less than 1, we conclude the series converges. Note: it is difficult to apply the Ratio Test to this series.
2. $\lim_{n\to\infty}\left(\frac{n^4}{(\ln n)^n}\right)^{1/n}=\lim_{n\to\infty}\frac{\left(n^{1/n}\right)^4}{\ln n}.$
   As n grows, the numerator approaches 1 (apply L'H\^opital's Rule) and the denominator grows to infinity. Thus
   $$\lim_{n\to\infty}\frac{\left(n^{1/n}\right)^4}{\ln n}=0. \tag{8.4.1}$$
   Since the limit is less than 1, we conclude the series converges.
3. $\lim_{n\to\infty}\left(\frac{2^n}{n^2}\right)^{1/n}=\lim_{n\to\infty}\frac{2}{\left(n^{1/n}\right)^2}=2.$
   Since this is greater than 1, we conclude the series diverges.

Each of the tests we have encountered so far has required that we analyze series from *positive* sequences. The next section relaxes this restriction by considering *alternating series*, where the underlying sequence has terms that alternate between being positive and negative.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 8.4: Ratio and Root Tests is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

8.5: Alternating Series and Absolute Convergence

All of the series convergence tests we have used require that the underlying sequence $\{a_n\}$ be a positive sequence. (We can relax this with Theorem 64 and state that there must be an $N>0$ such that $a_n>0$ for all $n>N$; that is, $\{a_n\}$ is positive for all but a finite number of values of n.)

In this section we explore series whose summation includes negative terms. We start with a very specific form of series, where the terms of the summation alternate between being positive and negative.

Definition 34: alternating series

Let $\{a_n\}$ be a positive sequence. An **alternating series** is a series of either the form

$$\sum_{n=1}^{\infty}(-1)^n a_n \qquad \text{or} \qquad \sum_{n=1}^{\infty}(-1)^{n+1}a_n. \tag{8.5.1}$$

Recall the terms of Harmonic Series come from the Harmonic Sequence $\{a_n\}=\{1/n\}$. An important alternating series is the **Alternating Harmonic Series**:

$$\sum_{n=1}^{\infty}(-1)^{n+1}\frac{1}{n}=1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}-\frac{1}{6}+\cdots \tag{8.5.2}$$

Geometric Series can also be alternating series when $r<0$. For instance, if $r=-1/2$, the geometric series is

$$\sum_{n=0}^{\infty}\left(\frac{-1}{2}\right)^n=1-\frac{1}{2}+\frac{1}{4}-\frac{1}{8}+\frac{1}{16}-\frac{1}{32}+\cdots \tag{8.5.3}$$

Theorem 60 states that geometric series converge when $|r|<1$ and gives the sum: $\sum_{n=0}^{\infty} r^n=\frac{1}{1-r}$. When $r=-1/2$ as above, we find

$$\sum_{n=0}^{\infty}\left(\frac{-1}{2}\right)^n=\frac{1}{1-(-1/2)}=\frac{1}{3/2}=\frac{2}{3}. \tag{8.5.4}$$

A powerful convergence theorem exists for other alternating series that meet a few conditions.

theorem 70: alternating series test

Let $\{a_n\}$ be a positive, decreasing sequence where $\lim_{n\to\infty} a_n=0$. Then

$$\sum_{n=1}^{\infty}(-1)^n a_n \qquad \text{and} \qquad \sum_{n=1}^{\infty}(-1)^{n+1}a_n \tag{8.5.5}$$

converge.

The basic idea behind Theorem 70 is illustrated in Figure 8.5.1. A positive, decreasing sequence $\{a_n\}$ is shown along with the partial sums

$$S_n=\sum_{i=1}^{n}(-1)^{i+1}a_i=a_1-a_2+a_3-a_4+\cdots+(-1)^{n+1}a_n. \tag{8.5.6}$$

Because $\{a_n\}$ is decreasing, the amount by which S_n bounces up/down decreases. Moreover, the odd terms of S_n form a decreasing, bounded sequence, while the even terms of S_n form an increasing, bounded sequence. Since bounded, monotonic sequences converge (see Theorem 59) and the terms of $\{a_n\}$ approach 0, one can show the odd and even terms of S_n converge to the same common limit L, the sum of the series.

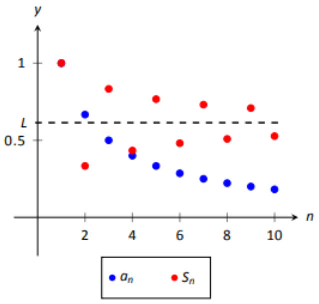


Figure 8.5.1: Illustrating the convergence with the Alternating Series Test.

Example 8.5.1: Applying the Alternating Series Test

Determine if the Alternating Series Test applies to each of the following series.

1. $\sum_{n=1}^{\infty} (-1)^{n+1} \frac{1}{n}$
2. $\sum_{n=1}^{\infty} (-1)^{n} \frac{\ln n}{n}$
3. $\sum_{n=1}^{\infty} (-1)^{n+1} \frac{|\sin n|}{n^2}$

Solution

1. This is the Alternating Harmonic Series as seen previously. The underlying sequence is $\{a_n\} = \{1/n\}$, which is positive, decreasing, and approaches 0 as $n \to \infty$. Therefore we can apply the Alternating Series Test and conclude this series converges.
   While the test does not state what the series converges to, we will see later that $\sum_{n=1}^{\infty} (-1)^{n+1} \frac{1}{n} = \ln 2$.
2. The underlying sequence is $\{a_n\} = \{\ln n/n\}$. This is positive and approaches 0 as $n \to \infty$ (use L'Hopital's Rule). However, the sequence is not decreasing for all n. It is straightforward to compute $a_1 = 0$, $a_2 \approx 0.347$, $a_3 \approx 0.366$, and $a_4 \approx 0.347$: the sequence is increasing for at least the first 3 terms.

   We do not immediately conclude that we cannot apply the Alternating Series Test. Rather, consider the long--term behavior of $\{a_n\}$. Treating $a_n = a(n)$ as a continuous function of n defined on $[1, \infty)$, we can take its derivative:

   $$a'(n) = \frac{1 - \ln n}{n^2}. \tag{8.5.7}$$

   The derivative is negative for all $n \geq 3$ (actually, for all $n > e$), meaning $a(n) = a_n$ is decreasing on $[3, \infty)$. We can apply the Alternating Series Test to the series when we start with $n = 3$ and conclude that $\sum_{n=3}^{\infty} (-1)^{n} \frac{\ln n}{n}$ converges; adding the terms with $n = 1$ and $n = 2$ do not change the convergence (i.e., we apply Theorem 64).

   The important lesson here is that as before, if a series fails to meet the criteria of the Alternating Series Test on only a finite number of terms, we can still apply the test.
3. The underlying sequence is $\{a_n\} = |\sin n|/n$. This sequence is positive and approaches 0 as $n \to \infty$. However, it is not a decreasing sequence; the value of $|\sin n|$ oscillates between 0 and 1 as $n \to \infty$. We cannot remove a finite number of terms to make $\{a_n\}$ decreasing, therefore we cannot apply the Alternating Series Test.

   Keep in mind that this does not mean we conclude the series diverges; in fact, it does converge. We are just unable to conclude this based on Theorem 70.

Key Idea 31 gives the sum of some important series. Two of these are

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} \approx 1.64493 \quad \text{and} \quad \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^2} = \frac{\pi^2}{12} \approx 0.82247. \tag{8.5.8}$$

These two series converge to their sums at different rates. To be accurate to two places after the decimal, we need 202 terms of the first series though only 13 of the second. To get 3 places of accuracy, we need 1069 terms of the first series though only 33 of the second. Why is it that the second series converges so much faster than the first?

While there are many factors involved when studying rates of convergence, the alternating structure of an alternating series gives us a powerful tool when approximating the sum of a convergent series.

theorem 71: the alternating series approximation theorem

Let $\{a_n\}$ be a sequence that satisfies the hypotheses of the Alternating Series Test, and let S_n and L be the n^{th} partial sums and sum, respectively, of either $\sum_{n=1}^{\infty}(-1)^n a_n$ or $\sum_{n=1}^{\infty}(-1)^{n+1}a_n$. Then

1. $|S_n - L| < a_{n+1}$, and
2. L is between S_n and S_{n+1}.

Part 1 of Theorem 71 states that the n^{th} partial sum of a convergent alternating series will be within a_{n+1} of its total sum. Consider the alternating series we looked at before the statement of the theorem, $\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}$. Since $a_{14} = 1/14^2 \approx 0.0051$, we know that S_{13} is within 0.0051 of the total sum.

Moreover, Part 2 of the theorem states that since $S_{13} \approx 0.8252$ and $S_{14} \approx 0.8201$, we know the sum L lies between 0.8201 and 0.8252. One use of this is the knowledge that S_{14} is accurate to two places after the decimal.

Some alternating series converge slowly. In Example $8.5.1$ we determined the series $\sum_{n=1}^{\infty}(-1)^{n+1}\frac{\ln n}{n}$ converged. With $n = 1001$, we find $\ln n/n \approx 0.0069$, meaning that $S_{1000} \approx 0.1633$ is accurate to one, maybe two, places after the decimal. Since $S_{1001} \approx 0.1564$, we know the sum L is $0.1564 \leq L \leq 0.1633$.

Example $8.5.2$: Approximating the sum of convergent alternating series

Approximate the sum of the following series, accurate to within 0.001.

1. $\sum_{n=1}^{\infty}(-1)^{n+1}\frac{1}{n^3}$ 2. $\sum_{n=1}^{\infty}(-1)^{n+1}\frac{\ln n}{n}$.

Solution

1. Using Theorem 71, we want to find n where $1/n^3 < 0.001$:

$$\begin{aligned}\frac{1}{n^3} &\leq 0.001 = \frac{1}{1000}\\ n^3 &\geq 1000\\ n &\geq \sqrt[3]{1000}\\ n &\geq 10.\end{aligned}$$

   Let L be the sum of this series. By Part 1 of the theorem, $|S_9 - L| < a_{10} = 1/1000$. We can compute $S_9 = 0.902116$, which our theorem states is within 0.001 of the total sum.

   We can use Part 2 of the theorem to obtain an even more accurate result. As we know the 10^{th} term of the series is $-1/1000$, we can easily compute $S_{10} = 0.901116$. Part 2 of the theorem states that L is between S_9 and S_{10}, so $0.901116 < L < 0.902116$.
2. We want to find n where $\ln(n)/n < 0.001$. We start by solving $\ln(n)/n = 0.001$ for n. This cannot be solved algebraically, so we will use Newton's Method to approximate a solution.

   Let $f(x) = \ln(x)/x - 0.001$; we want to know where $f(x) = 0$. We make a guess that x must be "large," so our initial guess will be $x_1 = 1000$. Recall how Newton's Method works: given an approximate solution x_n, our next approximation x_{n+1} is given by

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \tag{8.5.9}$$

We find $f'(x) = \big(1-\ln(x)\big)/x^2$. This gives

$$\begin{aligned} x_2 &= 1000 - \frac{\ln(1000)/1000 - 0.001}{\big(1-\ln(1000)\big)/1000^2} \\ &= 2000. \end{aligned}$$

Using a computer, we find that Newton's Method seems to converge to a solution $x = 9118.01$ after 8 iterations. Taking the next integer higher, we have $n = 9119$, where $\ln(9119)/9119 = 0.000999903 < 0.001$.

Again using a computer, we find $S_{9118} = -0.160369$. Part 1 of the theorem states that this is within 0.001 of the actual sum L. Already knowing the 9,119th term, we can compute $S_{9119} = -0.159369$, meaning $-0.159369 < L < -0.160369$.

Notice how the first series converged quite quickly, where we needed only 10 terms to reach the desired accuracy, whereas the second series took over 9,000 terms.

One of the famous results of mathematics is that the Harmonic Series, $\sum\limits_{n=1}^{\infty} \frac{1}{n}$ diverges, yet the Alternating Harmonic Series, $\sum\limits_{n=1}^{\infty} (-1)^{n+1}\frac{1}{n}$, converges. The notion that alternating the signs of the terms in a series can make a series converge leads us to the following definitions.

Definition 35: absolute and conditional convergence

1. A series $\sum\limits_{n=1}^{\infty} a_n$ **converges absolutely** if $\sum\limits_{n=1}^{\infty} |a_n|$ converges.
2. A series $\sum\limits_{n=1}^{\infty} a_n$ **converges conditionally** if $\sum\limits_{n=1}^{\infty} a_n$ converges but $\sum\limits_{n=1}^{\infty} |a_n|$ diverges.

Thus we say the Alternating Harmonic Series converges conditionally.

Example 8.5.3: Determining absolute and conditional convergence.

Determine if the following series converge absolutely, conditionally, or diverge.

1. $\sum\limits_{n=1}^{\infty} (-1)^n \frac{n+3}{n^2+2n+5}$ 2. $\sum\limits_{n=1}^{\infty} (-1)^n \frac{n^2+2n+5}{2^n}$ 3. $\sum\limits_{n=3}^{\infty} (-1)^n \frac{3n-3}{5n-10}$

Solution

1. We can show the series

$$\sum_{n=1}^{\infty} \left|(-1)^n \frac{n+3}{n^2+2n+5}\right| = \sum_{n=1}^{\infty} \frac{n+3}{n^2+2n+5} \tag{8.5.10}$$

   diverges using the Limit Comparison Test, comparing with $1/n$.

   The series $\sum\limits_{n=1}^{\infty} (-1)^n \frac{n+3}{n^2+2n+5}$ converges using the Alternating Series Test; we conclude it converges conditionally.
2. We can show the series

$$\sum_{n=1}^{\infty} \left|(-1)^n \frac{n^2+2n+5}{2^n}\right| = \sum_{n=1}^{\infty} \frac{n^2+2n+5}{2^n} \tag{8.5.11}$$

   converges using the Ratio Test.

   Therefore we conclude $\sum\limits_{n=1}^{\infty} (-1)^n \frac{n^2+2n+5}{2^n}$ converges absolutely.
3. The series

$$\sum_{n=3}^{\infty}\left|(-1)^n\frac{3n-3}{5n-10}\right| = \sum_{n=3}^{\infty}\frac{3n-3}{5n-10} \tag{8.5.12}$$

diverges using the n^{th} Term Test, so it does not converge absolutely.

The series $\sum_{n=3}^{\infty}(-1)^n\frac{3n-3}{5n-10}$ fails the conditions of the Alternating Series Test as $(3n-3)/(5n-10)$ does not approach 0 as $n\to\infty$. We can state further that this series diverges; as $n\to\infty$, the series effectively adds and subtracts $3/5$ over and over. This causes the sequence of partial sums to oscillate and not converge.

Therefore the series $\sum_{n=1}^{\infty}(-1)^n\frac{3n-3}{5n-10}$ diverges.

Knowing that a series converges absolutely allows us to make two important statements, given in the following theorem. The first is that absolute convergence is "stronger" than regular convergence. That is, just because $\sum_{n=1}^{\infty}a_n$ converges, we cannot conclude that $\sum_{n=1}^{\infty}|a_n|$ will converge, but knowing a series converges absolutely tells us that $\sum_{n=1}^{\infty}a_n$ will converge.

One reason this is important is that our convergence tests all require that the underlying sequence of terms be positive. By taking the absolute value of the terms of a series where not all terms are positive, we are often able to apply an appropriate test and determine absolute convergence. This, in turn, determines that the series we are given also converges.

The second statement relates to **rearrangements** of series. When dealing with a finite set of numbers, the sum of the numbers does not depend on the order which they are added. (So $1+2+3=3+1+2$.) One may be surprised to find out that when dealing with an infinite set of numbers, the same statement does not always hold true: some infinite lists of numbers may be rearranged in different orders to achieve different sums. The theorem states that the terms of an absolutely convergent series can be rearranged in any way without affecting the sum.

theorem 72: absolute convergence theorem

Let $\sum_{n=1}^{\infty}a_n$ be a series that converges absolutely.

1. $\sum_{n=1}^{\infty}a_n$ converges.
2. Let $\{b_n\}$ be any rearrangement of the sequence $\{a_n\}$. Then

$$\sum_{n=1}^{\infty}b_n = \sum_{n=1}^{\infty}a_n. \tag{8.5.13}$$

In Example 8.5.3, we determined the series in part 2 converges absolutely. Theorem 72 tells us the series converges (which we could also determine using the Alternating Series Test).

The theorem states that rearranging the terms of an absolutely convergent series does not affect its sum. This implies that perhaps the sum of a conditionally convergent series can change based on the arrangement of terms. Indeed, it can. The Riemann Rearrangement Theorem (named after Bernhard Riemann) states that any conditionally convergent series can have its terms rearranged so that the sum is any desired value, including ∞!

As an example, consider the Alternating Harmonic Series once more. We have stated that

$$\sum_{n=1}^{\infty}(-1)^{n+1}\frac{1}{n} = 1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}-\frac{1}{6}+\frac{1}{7}\cdots = \ln 2, \tag{8.5.14}$$

(see Key Idea 31 or Example 8.5.1).

Consider the rearrangement where every positive term is followed by two negative terms:

$$1-\frac{1}{2}-\frac{1}{4}+\frac{1}{3}-\frac{1}{6}-\frac{1}{8}+\frac{1}{5}-\frac{1}{10}-\frac{1}{12}\cdots \tag{8.5.15}$$

(Convince yourself that these are exactly the same numbers as appear in the Alternating Harmonic Series, just in a different order.) Now group some terms and simplify:

$$\begin{aligned}\left(1-\frac{1}{2}\right)-\frac{1}{4}+\left(\frac{1}{3}-\frac{1}{6}\right)-\frac{1}{8}+\left(\frac{1}{5}-\frac{1}{10}\right)-\frac{1}{12}+\cdots &= \\ \frac{1}{2}-\frac{1}{4}+\frac{1}{6}-\frac{1}{8}+\frac{1}{10}-\frac{1}{12}+\cdots &= \\ \frac{1}{2}\left(1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}-\frac{1}{6}+\cdots\right) &= \frac{1}{2}\ln 2.\end{aligned}$$

By rearranging the terms of the series, we have arrived at a different sum! (One could *try* to argue that the Alternating Harmonic Series does not actually converge to $\ln 2$, because rearranging the terms of the series *shouldn't* change the sum. However, the Alternating Series Test proves this series converges to L, for some number L, and if the rearrangement does not change the sum, then $L=L/2$, implying $L=0$. But the Alternating Series Approximation Theorem quickly shows that $L>0$. The only conclusion is that the rearrangement \emph{did} change the sum.) This is an incredible result.

We end here our study of tests to determine convergence. The back cover of this text contains a table summarizing the tests that one may find useful.

While series are worthy of study in and of themselves, our ultimate goal within calculus is the study of Power Series, which we will consider in the next section. We will use power series to create functions where the output is the result of an infinite summation.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 8.5: Alternating Series and Absolute Convergence is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

8.6: Power Series

So far, our study of series has examined the question of "Is the sum of these infinite terms finite?," i.e., "Does the series converge?" We now approach series from a different perspective: as a function. Given a value of x, we evaluate $f(x)$ by finding the sum of a particular series that depends on x (assuming the series converges). We start this new approach to series with a definition.

Definition 36: power series

Let $\{a_n\}$ be a sequence, let x be a variable, and let c be a real number.

1. The **power series in** x is the series

$$\sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots \tag{8.6.1}$$

2. The **power series in** x centered at c is the series

$$\sum_{n=0}^{\infty} a_n (x-c)^n = a_0 + a_1(x-c) + a_2(x-c)^2 + a_3(x-c)^3 + \dots \tag{8.6.2}$$

Example 8.6.1: Examples of power series

Write out the first five terms of the following power series:

1. $\sum\limits_{n=0}^{\infty} x^n$ 2. $\sum\limits_{n=1}^{\infty} (-1)^{n+1} \frac{(x+1)^n}{n}$ 3. $\sum\limits_{n=0}^{\infty} (-1)^{n+1} \frac{(x-\pi)^{2n}}{(2n)!}$.

Solution

1. One of the conventions we adopt is that $x^0 = 1$ regardless of the value of x. Therefore

$$\sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + x^4 + \dots \tag{8.6.3}$$

This is a geometric series in x.

2. This series is centered at $c = -1$. Note how this series starts with $n = 1$. We could rewrite this series starting at $n = 0$ with the understanding that $a_0 = 0$, and hence the first term is 0.

$$\begin{aligned} &\sum_{n=1}^{\infty} (-1)^{n+1} \frac{(x+1)^n}{n} \\ &= (x+1) - \frac{(x+1)^2}{2} + \frac{(x+1)^3}{3} - \frac{(x+1)^4}{4} + \frac{(x+1)^5}{5} \dots \end{aligned} \tag{8.6.4}$$

3. This series is centered at $c = \pi$. Recall that $0! = 1$.

$$\sum_{n=0}^{\infty} (-1)^{n+1} \frac{(x-\pi)^{2n}}{(2n)!} = -1 + \frac{(x-\pi)^2}{2} - \frac{(x-\pi)^4}{24} + \frac{(x-\pi)^6}{6!} - \frac{(x-\pi)^8}{8!} \dots \tag{8.6.5}$$

We introduced power series as a type of function, where a value of x is given and the sum of a series is returned. Of course, not every series converges. For instance, in part 1 of Example 8.6.1, we recognized the series $\sum\limits_{n=0}^{\infty} x^n$ as a geometric series in x. Theorem 60 states that this series converges only when $|x| < 1$.

This raises the question: "For what values of x will a given power series converge?," which leads us to a theorem and definition.

theorem 73: convergence of power series

Let a power series $\sum\limits_{n=0}^{\infty} a_n (x-c)^n$ be given. Then one of the following is true:

1. The series converges only at $x = c$.
2. There is an $R > 0$ such that the series converges for all x in $(c-R, c+R)$ and diverges for all $x < c - R$ and $x > c + R$.
3. The series converges for all x.

The value of R is important when understanding a power series, hence it is given a name in the following definition. Also, note that part 2 of Theorem 73 makes a statement about the interval $(c-R, c+R)$, but the not the endpoints of that interval. A series may/may not converge at these endpoints.

Definition 37: Radius and Interval of Convergence

1. The number R given in Theorem 73 is the **radius of convergence** of a given series. When a series converges for only $x=c$, we say the radius of convergence is 0, i.e., $R=0$. When a series converges for all x, we say the series has an infinite radius of convergence, i.e., $R=\infty$.
2. The **interval of convergence** is the set of all values of x for which the series converges.

To find the values of x for which a given series converges, we will use the convergence tests we studied previously (especially the Ratio Test). However, the tests all required that the terms of a series be positive. The following theorem gives us a work--around to this problem.

theorem 74: The Radius of Convergence of a Series and Absolute Convergence

The series $\sum\limits_{n=0}^{\infty} a_n x^n$ and $\sum\limits_{n=0}^{\infty} |a_n x^n|$ have the same radius of convergence R.

Theorem 74 allows us to find the radius of convergence R of a series by applying the Ratio Test (or any applicable test) to the absolute value of the terms of the series. We practice this in the following example.

Example $8.6.2$: Determining the radius and interval of convergence.

Find the radius and interval of convergence for each of the following series:

1. $\sum\limits_{n=0}^{\infty} \frac{x^n}{n!}$
2. $\sum\limits_{n=1}^{\infty} (-1)^{n+1}\frac{x^n}{n}$
3. $\sum\limits_{n=0}^{\infty} 2^n(x-3)^n$
4. $\sum\limits_{n=0}^{\infty} n!x^n$

Solution

1. We apply the Ratio Test to the series $\sum\limits_{n=0}^{\infty} \left|\frac{x^n}{n!}\right|$:

$$\begin{aligned}\lim_{n\to\infty}\frac{|x^{n+1}/(n+1)!|}{|x^n/n!|} &= \lim_{n\to\infty}\left|\frac{x^{n+1}}{x^n}\cdot\frac{n!}{(n+1)!}\right| \\ &= \lim_{n\to\infty}\left|\frac{x}{n+1}\right| \\ &= 0 \text{ for all } x.\end{aligned}$$

2. The Ratio Test shows us that regardless of the choice of x, the series converges. Therefore the radius of convergence is $R = \infty$, and the interval of convergence is $(-\infty, \infty)$.
3. We apply the Ratio Test to the series $\sum\limits_{n=1}^{\infty} \left|(-1)^{n+1}\frac{x^n}{n}\right| = \sum\limits_{n=1}^{\infty}\left|\frac{x^n}{n}\right|$:

$$\begin{aligned}\lim_{n\to\infty}\frac{|x^{n+1}/(n+1)|}{|x^n/n|} &= \lim_{n\to\infty}\left|\frac{x^{n+1}}{x^n}\cdot\frac{n}{n+1}\right| \\ &= \lim_{n\to\infty}|x|\frac{n}{n+1} \\ &= |x|.\end{aligned}$$

The Ratio Test states a series converges if the limit of $|a_{n+1}/a_n| = L < 1$. We found the limit above to be $|x|$; therefore, the power series converges when $|x| < 1$, or when x is in $(-1, 1)$. Thus the radius of convergence is $R = 1$.

To determine the interval of convergence, we need to check the endpoints of $(-1,1)$. When $x=-1$, we have the opposite of the Harmonic Series:

$$\begin{aligned}\sum_{n=1}^{\infty}(-1)^{n+1}\frac{(-1)^n}{n} &= \sum_{n=1}^{\infty}\frac{-1}{n}\\ &= -\infty.\end{aligned}$$

The series diverges when $x=-1$.

When $x=1$, we have the series $\sum_{n=1}^{\infty}(-1)^{n+1}\frac{(1)^n}{n}$, which is the Alternating Harmonic Series, which converges. Therefore the interval of convergence is $(-1,1]$.

4. We apply the Ratio Test to the series $\sum_{n=0}^{\infty}\left|2^n(x-3)^n\right|$:

$$\begin{aligned}\lim_{n\to\infty}\frac{\left|2^{n+1}(x-3)^{n+1}\right|}{\left|2^n(x-3)^n\right|} &= \lim_{n\to\infty}\left|\frac{2^{n+1}}{2^n}\cdot\frac{(x-3)^{n+1}}{(x-3)^n}\right|\\ &= \lim_{n\to\infty}\left|2(x-3)\right|.\end{aligned}$$

According to the Ratio Test, the series converges when $\left|2(x-3)\right|<1 \implies |x-3|<1/2$. The series is centered at 3, and x must be within $1/2$ of 3 in order for the series to converge. Therefore the radius of convergence is $R=1/2$, and we know that the series converges absolutely for all x in $(3-1/2,3+1/2)=(2.5,3.5)$.

We check for convergence at the endpoints to find the interval of convergence. When $x=2.5$, we have:

$$\begin{aligned}\sum_{n=0}^{\infty}2^n(2.5-3)^n &= \sum_{n=0}^{\infty}2^n(-1/2)^n\\ &= \sum_{n=0}^{\infty}(-1)^n,\end{aligned}$$

which diverges. A similar process shows that the series also diverges at $x=3.5$. Therefore the interval of convergence is $(2.5,3.5)$.

5. We apply the Ratio Test to $\sum_{n=0}^{\infty}\left|n!x^n\right|$:

$$\begin{aligned}\lim_{n\to\infty}\frac{\left|(n+1)!x^{n+1}\right|}{\left|n!x^n\right|} &= \lim_{n\to\infty}\left|(n+1)x\right|\\ &= \infty \quad \text{for all } x, \text{ except } x=0.\end{aligned}$$

The Ratio Test shows that the series diverges for all x except $x=0$. Therefore the radius of convergence is $R=0$.

We can use a power series to define a function:

$$f(x)=\sum_{n=0}^{\infty}a_nx^n \tag{8.6.6}$$

where the domain of f is a subset of the interval of convergence of the power series. One can apply calculus techniques to such functions; in particular, we can find derivatives and antiderivatives.

theorem 75: Derivatives and Indefinite Integrals of Power Series Functions

Let $f(x)=\sum_{n=0}^{\infty}a_n(x-c)^n$ be a function defined by a power series, with radius of convergence R.

1. $f(x)$ is continuous and differentiable on $(c-R,c+R)$.
2. $f'(x)=\sum_{n=1}^{\infty}a_n\cdot n\cdot(x-c)^{n-1}$, with radius of convergence R.

3. $\int f(x)\ dx = C + \sum\limits_{n=0}^{\infty} a_n \frac{(x-c)^{n+1}}{n+1}$, with radius of convergence R.

A few notes about Theorem 75:

1. The theorem states that differentiation and integration do not change the radius of convergence. It does not state anything about the *interval* of convergence. They are not always the same.
2. Notice how the summation for $f'(x)$ starts with $n = 1$. This is because the constant term a_0 of $f(x)$ goes to 0.
3. Differentiation and integration are simply calculated term--by--term using the Power Rules.

Example 8.6.3: Derivatives and indefinite integrals of power series

Let $f(x) = \sum\limits_{n=0}^{\infty} x^n$. Find $f'(x)$ and $F(x) = \int f(x)\ dx$, along with their respective intervals of convergence.

Solution

We find the derivative and indefinite integral of $f(x)$, following Theorem 75.

1. $f'(x) = \sum\limits_{n=1}^{\infty} nx^{n-1} = 1 + 2x + 3x^2 + 4x^3 + \cdots$.

   In Example 8.6.1, we recognized that $\sum\limits_{n=0}^{\infty} x^n$ is a geometric series in x. We know that such a geometric series converges when $|x| < 1$; that is, the interval of convergence is $(-1, 1)$.

   To determine the interval of convergence of $f'(x)$, we consider the endpoints of $(-1, 1)$:

   $$f'(-1) = 1 - 2 + 3 - 4 + \cdots, \quad \text{which diverges.} \tag{8.6.7}$$

   $$f'(1) = 1 + 2 + 3 + 4 + \cdots, \quad \text{which diverges.} \tag{8.6.8}$$

   Therefore, the interval of convergence of $f'(x)$ is $(-1, 1)$.
2. $F(x) = \int f(x)\ dx = C + \sum\limits_{n=0}^{\infty} \frac{x^{n+1}}{n+1} = C + x + \frac{x^2}{2} + \frac{x^3}{3} + \cdots$

   To find the interval of convergence of $F(x)$, we again consider the endpoints of $(-1, 1)$:

   $$F(-1) = C - 1 + 1/2 - 1/3 + 1/4 + \cdots \tag{8.6.9}$$

   The value of C is irrelevant; notice that the rest of the series is an Alternating Series that whose terms converge to 0. By the Alternating Series Test, this series converges. (In fact, we can recognize that the terms of the series after C are the opposite of the Alternating Harmonic Series. We can thus say that $F(-1) = C - \ln 2$.)

   $$F(1) = C + 1 + 1/2 + 1/3 + 1/4 + \cdots \tag{8.6.10}$$

   Notice that this summation is C + the Harmonic Series, which diverges. Since F converges for $x = -1$ and diverges for $x = 1$, the interval of convergence of $F(x)$ is $[-1, 1)$.

The previous example showed how to take the derivative and indefinite integral of a power series without motivation for why we care about such operations. We may care for the sheer mathematical enjoyment "that we can", which is motivation enough for many. However, we would be remiss to not recognize that we can learn a great deal from taking derivatives and indefinite integrals.

Recall that $f(x) = \sum\limits_{n=0}^{\infty} x^n$ in Example 8.6.3 is a geometric series. According to Theorem 60, this series converges to $1/(1-x)$ when $|x| < 1$. Thus we can say

$$f(x) = \sum^{\infty} x^n = \frac{1}{\quad} \quad \text{on} \quad (-1, 1) \tag{8.6.11}$$

$$f(x)=\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}, \quad \text{on} \quad (-1,1). \tag{8.6.11}$$

Integrating the power series, (as done in Example 8.6.3,) we find

$$F(x)=C_1+\sum_{n=0}^{\infty}\frac{x^{n+1}}{n+1}, \tag{8.6.12}$$

while integrating the function $f(x)=1/(1-x)$ gives

$$F(x)=-\ln|1-x|+C_2. \tag{8.6.13}$$

Equating Equations 8.6.12 and 8.6.13, we have

$$F(x)=C_1+\sum_{n=0}^{\infty}\frac{x^{n+1}}{n+1}=-\ln|1-x|+C_2. \tag{8.6.14}$$

Letting $x=0$, we have $F(0)=C_1=C_2$. This implies that we can drop the constants and conclude

$$\sum_{n=0}^{\infty}\frac{x^{n+1}}{n+1}=-\ln|1-x|. \tag{8.6.15}$$

We established in Example 8.6.3 that the series on the left converges at $x=-1$; substituting $x=-1$ on both sides of the above equality gives

$$-1+\frac{1}{2}-\frac{1}{3}+\frac{1}{4}-\frac{1}{5}+\cdots=-\ln 2. \tag{8.6.16}$$

On the left we have the opposite of the Alternating Harmonic Series; on the right, we have $-\ln 2$. We conclude that

$$1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\cdots=\ln 2. \tag{8.6.17}$$

Important: We stated in Key Idea 31 (in Section 8.2) that the Alternating Harmonic Series converges to $\ln 2$, and referred to this fact again in Section 8.5. However, we never gave an argument for why this was the case. The work above finally shows how we conclude that the Alternating Harmonic Series converges to $\ln 2$.

We use this type of analysis in the next example.

Example 8.6.4: Analyzing power series functions

Let $f(x)=\sum_{n=0}^{\infty}\frac{x^n}{n!}$. Find $f'(x)$ and $\int f(x)\,dx$, and use these to analyze the behavior of $f(x)$.

Solution

We start by making two notes: first, in Example 8.6.2, we found the interval of convergence of this power series is $(-\infty,\infty)$. Second, we will find it useful later to have a few terms of the series written out:

$$\sum_{n=0}^{\infty}\frac{x^n}{n!}=1+x+\frac{x^2}{2}+\frac{x^3}{6}+\frac{x^4}{24}+\cdots \tag{8.6.18}$$

We now find the derivative:

$$\begin{aligned} f'(x) &= \sum_{n=1}^{\infty} n\frac{x^{n-1}}{n!} \\ &= \sum_{n=1}^{\infty}\frac{x^{n-1}}{(n-1)!} = 1+x+\frac{x^2}{2!}+\cdots. \end{aligned}$$

Since the series starts at $n=1$ and each term refers to $(n-1)$, we can re-index the series starting with $n=0$:

$$\begin{aligned} &= \sum_{n=0}^{\infty}\frac{x^n}{n!} \\ &= f(x). \end{aligned}$$

We found the derivative of $f(x)$ is $f(x)$. The only functions for which this is true are of the form $y=ce^x$ for some constant c. As $f(0)=1$ (see Equation 8.6.18), c must be 1. Therefore we conclude that

$$f(x)=\sum_{n=0}^{\infty}\frac{x^n}{n!}=e^x \tag{8.6.19}$$

for all x.

We can also find $\int f(x)\ dx$:

$$\begin{aligned}\int f(x)dx &= C+\sum_{n=0}^{\infty}\frac{x^{n+1}}{n!(n+1)}\\ &= C+\sum_{n=0}^{\infty}\frac{x^{n+1}}{(n+1)!}\end{aligned}$$

We write out a few terms of this last series:

$$C+\sum_{n=0}^{\infty}\frac{x^{n+1}}{(n+1)!}=C+x+\frac{x^2}{2}+\frac{x^3}{6}+\frac{x^4}{24}+\cdots \tag{8.6.20}$$

The integral of $f(x)$ differs from $f(x)$ only by a constant, again indicating that $f(x)=e^x$.

Example 8.6.4 and the work following Example 8.6.3 established relationships between a power series function and "regular" functions that we have dealt with in the past. In general, given a power series function, it is difficult (if not impossible) to express the function in terms of elementary functions. We chose examples where things worked out nicely.

In this section's last example, we show how to solve a simple differential equation with a power series.\

Example $8.6.5$: Solving a differential equation with a power series.

Give the first 4 terms of the power series solution to $y'=2y$, where $y(0)=1$.

Solution

The differential equation $y'=2y$ describes a function $y=f(x)$ where the derivative of y is twice y and $y(0)=1$. This is a rather simple differential equation; with a bit of thought one should realize that if $y=Ce^{2x}$, then $y'=2Ce^{2x}$, and hence $y'=2y$. By letting $C=1$ we satisfy the initial condition of $y(0)=1$.

Let's ignore the fact that we already know the solution and find a power series function that satisfies the equation. The solution we seek will have the form

$$f(x)=\sum_{n=0}^{\infty}a_nx^n=a_0+a_1x+a_2x^2+a_3x^3+\cdots \tag{8.6.21}$$

for unknown coefficients a_n. We can find $f'(x)$ using Theorem 75:

$$f'(x)=\sum_{n=1}^{\infty}a_n\cdot n\cdot x^{n-1}=a_1+2a_2x+3a_3x^2+4a_4x^3\cdots. \tag{8.6.22}$$

Since $f'(x)=2f(x)$, we have

$$\begin{aligned}a_1+2a_2x+3a_3x^2+4a_4x^3\cdots &= 2\big(a_0+a_1x+a_2x^2+a_3x^3+\cdots\big)\\ &= 2a_0+2a_1x+2a_2x^2+2a_3x^3+\cdots\end{aligned}$$

The coefficients of like powers of x must be equal, so we find that

$$a_1=2a_0,\quad 2a_2=2a_1,\quad 3a_3=2a_2,\quad 4a_4=2a_3,\quad \text{etc.} \tag{8.6.23}$$

The initial condition $y(0)=f(0)=1$ indicates that $a_0=1$; with this, we can find the values of the other coefficients:

$$\begin{gathered}a_0=1\text{ and }a_1=2a_0\Rightarrow a_1=2;\\ a_1=2\text{ and }2a_2=2a_1\Rightarrow a_2=4/2=2;\\ a_2=2\text{ and }3a_3=2a_2\Rightarrow a_3=8/(2\cdot 3)=4/3;\\ a_3=4/3\text{ and }4a_4=2a_3\Rightarrow a_4=16/(2\cdot 3\cdot 4)=2/3.\end{gathered}$$

Thus the first 5 terms of the power series solution to the differential equation $y'=2y$ is

$$f(x)=1+2x+2x^2+\frac{4}{3}x^3+\frac{2}{3}x^4+\cdots \tag{8.6.24}$$

In Section 8.8, as we study Taylor Series, we will learn how to recognize this series as describing $y=e^{2x}$.

Our last example illustrates that it can be difficult to recognize an elementary function by its power series expansion. It is far easier to start with a known function, expressed in terms of elementary functions, and represent it as a power series function. One may wonder why we would bother doing so, as the latter function probably seems more complicated. In the next two sections, we show both *how* to do this and *why* such a process can be beneficial.

Contributors and Attributions

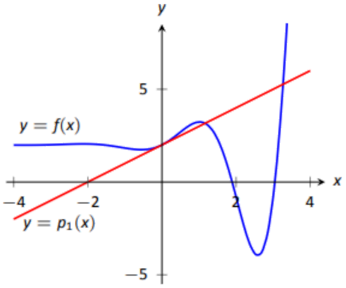
- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 8.6: Power Series is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

8.7: Taylor Polynomials

Consider a function $y = f(x)$ and a point $(c, f(c))$. The derivative, $f'(c)$, gives the instantaneous rate of change of f at $x = c$. Of all lines that pass through the point $(c, f(c))$, the line that best approximates f at this point is the tangent line; that is, the line whose slope (rate of change) is $f'(c)$.

In Figure 8.7.1, we see a function $y = f(x)$ graphed. The table below the graph shows that $f(0) = 2$ and $f'(0) = 1$; therefore, the tangent line to f at $x = 0$ is $p_1(x) = 1(x - 0) + 2 = x + 2$. The tangent line is also given in the figure. Note that "near" $x = 0$, $p_1(x) \approx f(x)$; that is, the tangent line approximates f well.



$f(0) = 2$ $\quad f'''(0) = -1$

$f'(0) = 1$ $\quad f^{(4)}(0) = -12$

$f''(0) = 2$ $\quad f^{(5)}(0) = -19$

Figure 8.7.1: Plotting $y = f(x)$ and a table of derivatives of f evaluated at 0.

One shortcoming of this approximation is that the tangent line only matches the slope of f; it does not, for instance, match the concavity of f. We can find a polynomial, $p_2(x)$, that does match the concavity without much difficulty, though. The table in Figure 8.16 gives the following information:

$$f(0) = 2 \qquad f'(0) = 1 \qquad f''(0) = 2. \tag{8.7.1}$$

Therefore, we want our polynomial $p_2(x)$ to have these same properties. That is, we need

$$p_2(0) = 2 \qquad p_2'(0) = 1 \qquad p_2''(0) = 2. \tag{8.7.2}$$

This is simply an initial-value problem. We can solve this using the techniques first described in Section 5.1. To keep $p_2(x)$ as simple as possible, we'll assume that not only $p_2''(0) = 2$, but that $p_2''(x) = 2$. That is, the second derivative of p_2 is constant.

If $p_2''(x) = 2$, then $p_2'(x) = 2x + C$ for some constant C. Since we have determined that $p_2'(0) = 1$, we find that $C = 1$ and so $p_2'(x) = 2x + 1$. Finally, we can compute $p_2(x) = x^2 + x + C$. Using our initial values, we know $p_2(0) = 2$ so $C = 2$. We conclude that $p_2(x) = x^2 + x + 2$. This function is plotted with f in Figure 8.7.2.

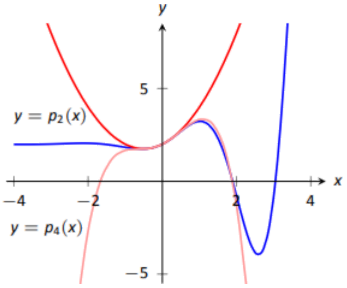


Figure 8.7.2: Plotting f (blue), p_2 (red) and p_4 (light red).

We can repeat this approximation process by creating polynomials of higher degree that match more of the derivatives of f at $x = 0$. In general, a polynomial of degree n can be created to match the first n derivatives of f. Figure 8.7.2 also shows $p_4(x) = -x^4/2 - x^3/6 + x^2 + x + 2$, whose first four derivatives at 0 match those of f. (Using the table in Figure 8.7.1, start with $p_4^{(4)}(x) = -12$ and solve the related initial-value problem.)

As we use more and more derivatives, our polynomial approximation to f gets better and better. In this example, the interval on which the approximation is "good" gets bigger and bigger. Figure 8.7.3 shows $p_{13}(x)$; we can visually affirm that this polynomial approximates f very well on $[-2, 3]$.

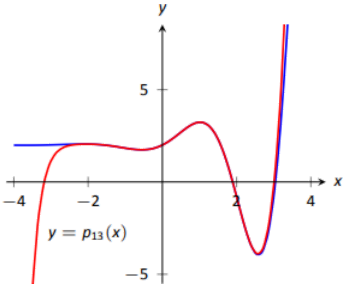


Figure 8.7.3: Plotting f and p_{13}.

The polynomial $p_{13}(x)$ is not particularly "nice". It is

$$p_{13} = \frac{16901x^{13}}{6227020800} + \frac{13x^{12}}{1209600} - \frac{1321x^{11}}{39916800} - \frac{779x^{10}}{1814400} - \frac{359x^{9}}{362880} + \frac{x^8}{240} + \frac{139x^7}{5040} + \frac{11x^6}{360} - \frac{19x^5}{120} - \frac{x^4}{2} - \frac{x^3}{6} + x^2 + x + 2. \tag{8.7.3}$$

The polynomials we have created are examples of *Taylor polynomials*, named after the British mathematician Brook Taylor who made important discoveries about such functions. While we created the above Taylor polynomials by solving initial-value problems, it can be shown that Taylor polynomials follow a general pattern that make their formation much more direct. This is described in the following definition.

Definition 38: Taylor Polynomials & Maclaurin PolynomiaLS

Let f be a function whose first n derivatives exist at $x = c$.

1. The Taylor polynomial of degree n of f at $x = c$ is

$$p_n(x) = f(c) + f'(c)(x-c) + \frac{f''(c)}{2!}(x-c)^2 + \frac{f'''(c)}{3!}(x-c)^3 + \cdots + \frac{f^{(n)}(c)}{n!}(x-c)^n. \tag{8.7.4}$$

2. A special case of the Taylor polynomial is the Maclaurin polynomial, where $c = 0$. That is, the **Maclaurin polynomial of degree** n of f is

$$p_n(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \cdots + \frac{f^{(n)}(0)}{n!}x^n. \tag{8.7.5}$$

We will practice creating Taylor and Maclaurin polynomials in the following examples.

Example 8.7.1: Finding and using Maclaurin polynomials

1. Find the n^{th} Maclaurin polynomial for $f(x) = e^x$.
2. Use $p_5(x)$ to approximate the value of e.

Solution

$$\begin{array}{lcl} f(x) = e^x & \Rightarrow & f(0) = 1 \\ f'(x) = e^x & \Rightarrow & f'(0) = 1 \\ f''(x) = e^x & \Rightarrow & f''(0) = 1 \\ \vdots & & \vdots \\ f^{(n)}(x) = e^x & \Rightarrow & f^{(n)}(0) = 1 \end{array}$$

Figure 8.7.4 The derivatives of $f(x) = e^x$ evaluated at $x = 0$.

1. We start with creating a table of the derivatives of e^x evaluated at $x = 0$. In this particular case, this is relatively simple, as shown in Figure 8.19.

   By the definition of the Maclaurin series, we have

$$\begin{aligned} p_n(x) &= f(0)+f'(0)x+\frac{f''(0)}{2!}x^2+\frac{f'''(0)}{3!}x^3+\cdots+\frac{f^{\,n}(0)}{n!}x^n \\ &= 1+x+\frac{1}{2}x^2+\frac{1}{6}x^3+\frac{1}{24}x^4+\cdots+\frac{1}{n!}x^n. \end{aligned}$$

2. Using our answer from part 1, we have

$$p_5=1+x+\frac{1}{2}x^2+\frac{1}{6}x^3+\frac{1}{24}x^4+\frac{1}{120}x^5. \tag{8.7.6}$$

To approximate the value of e, note that $e=e^1=f(1)\approx p_5(1)$. It is very straightforward to evaluate $p_5(1)$:

$$p_5(1)=1+1+\frac{1}{2}+\frac{1}{6}+\frac{1}{24}+\frac{1}{120}=\frac{163}{60}\approx 2.71667. \tag{8.7.7}$$

A plot of $f(x)=e^x$ and $p_5(x)$ is given in Figure $8.7.5$.

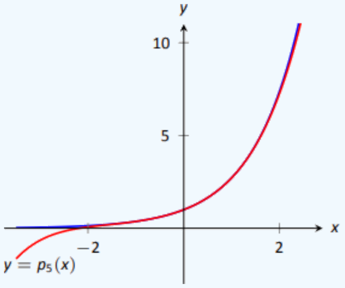


Figure $8.7.5$: A plot of $f(x)=e^x$ and its 5th degree Maclaurin polynomial $p_5(x)$.

Example $8.7.2$: Finding and using Taylor polynomials

1. Find the n^{th} Taylor polynomial of $y=\ln x$ at $x=1$.
2. Use $p_6(x)$ to approximate the value of $\ln 1.5$.
3. Use $p_6(x)$ to approximate the value of $\ln 2$.

Solution

$$\begin{array}{lcl} f(x)=\ln x & \Rightarrow & f(1)=0 \\ f'(x)=1/x & \Rightarrow & f'(1)=1 \\ f''(x)=-1/x^2 & \Rightarrow & f''(1)=-1 \\ f'''(x)=2/x^3 & \Rightarrow & f'''(1)=2 \\ f^{(4)}(x)=-6/x^4 & \Rightarrow & f^{(4)}(1)=-6 \\ \vdots & & \vdots \\ f^{(n)}(x)= & \Rightarrow & f^{(n)}(1)= \\ \dfrac{(-1)^{n+1}(n-1)!}{x^n} & & (-1)^{n+1}(n-1)! \end{array}$$

Figure $8.7.6$: Derivatives of $\ln x$ evaluated at $x=1$.

1. We begin by creating a table of derivatives of $\ln x$ evaluated at $x=1$. While this is not as straightforward as it was in the previous example, a pattern does emerge, as shown in Figure $8.7.6$.
   Using Definition 38, we have

$$\begin{aligned} p_n(x) &= f(c)+f'(c)(x-c)+\frac{f''(c)}{2!}(x-c)^2+\frac{f'''(c)}{3!}(x-c)^3+\cdots+\frac{f^{\,n}(c)}{n!}(x-c)^n \\ &= 0+(x-1)-\frac{1}{2}(x-1)^2+\frac{1}{3}(x-1)^3-\frac{1}{4}(x-1)^4+\cdots+\frac{(-1)^{n+1}}{n}(x-1)^n. \end{aligned}$$

   Note how the coefficients of the $(x-1)$ terms turn out to be "nice."
2. We can compute $p_6(x)$ using our work above:

$$\begin{aligned} &p_6(x) \\ &=(x-1)-\frac{1}{2}(x-1)^2+\frac{1}{3}(x-1)^3-\frac{1}{4}(x-1)^4+\frac{1}{5}(x-1)^5-\frac{1}{6}(x-1)^6. \end{aligned} \tag{8.7.8}$$

Since $p_6(x)$ approximates $\ln x$ well near $x=1$, we approximate $\ln 1.5 \approx p_6(1.5)$:

$$\begin{aligned} p_6(1.5) &= (1.5-1) - \frac{1}{2}(1.5-1)^2 + \frac{1}{3}(1.5-1)^3 - \frac{1}{4}(1.5-1)^4 + \cdots \\ &\cdots + \frac{1}{5}(1.5-1)^5 - \frac{1}{6}(1.5-1)^6 \\ &= \frac{259}{640} \\ &\approx 0.404688. \end{aligned}$$

This is a good approximation as a calculator shows that $\ln 1.5 \approx 0.4055$. Figure $8.7.7$ plots $y=\ln x$ with $y=p_6(x)$. We can see that $\ln 1.5 \approx p_6(1.5)$.

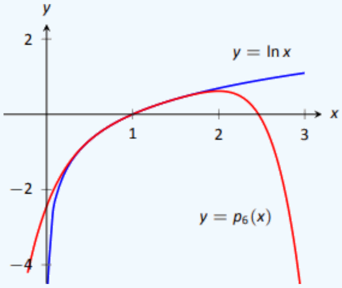


Figure $8.7.7$: A plot of $y=\ln x$ and its 6th degree Taylor polynomial at $x=1$.

3. We approximate $\ln 2$ with $p_6(2)$:

$$\begin{aligned} p_6(2) &= (2-1) - \frac{1}{2}(2-1)^2 + \frac{1}{3}(2-1)^3 - \frac{1}{4}(2-1)^4 + \cdots \\ &\cdots + \frac{1}{5}(2-1)^5 - \frac{1}{6}(2-1)^6 \\ &= 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} \\ &= \frac{37}{60} \\ &\approx 0.616667. \end{aligned}$$

This approximation is not terribly impressive: a hand held calculator shows that $\ln 2 \approx 0.693147$. The graph in Figure 8.22 shows that $p_6(x)$ provides less accurate approximations of $\ln x$ as x gets close to 0 or 2.

Surprisingly enough, even the 20$^{\text{th}}$ degree Taylor polynomial fails to approximate $\ln x$ for $x>2$, as shown in Figure $8.7.8$. We'll soon discuss why this is.

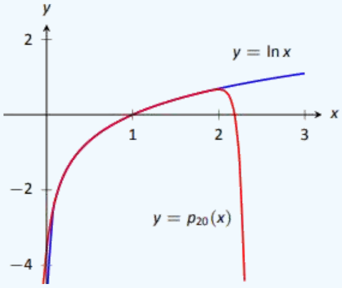


Figure $8.7.8$: A plot of $y=\ln x$ and its 20th degree Taylor polynomial at $x=1$.

Taylor polynomials are used to approximate functions $f(x)$ in mainly two situations:

1. When $f(x)$ is known, but perhaps "hard" to compute directly. For instance, we can define $y=\cos x$ as either the ratio of sides of a right triangle ("adjacent over hypotenuse") or with the unit circle. However, neither of these provides a convenient way of computing $\cos 2$. A Taylor polynomial of sufficiently high degree can provide a reasonable method of computing such values using only operations usually hard-wired into a computer ($+$, $-$, $\times$ and $\div$).
2. When $f(x)$ is not known, but information about its derivatives is known. This occurs more often than one might think, especially in the study of differential equations.

> Even though Taylor polynomials *could* be used in calculators and computers to calculate values of trigonometric functions, in practice they generally aren't. Other more efficient and accurate methods have been developed, such as the CORDIC algorithm.

In both situations, a critical piece of information to have is "How good is my approximation?" If we use a Taylor polynomial to compute $\cos 2$, how do we know how accurate the approximation is?

We had the same problem when studying Numerical Integration. Theorem 43 provided bounds on the error when using, say, Simpson's Rule to approximate a definite integral. These bounds allowed us to determine that, for instance, using 10 subintervals provided an approximation within $\pm .01$ of the exact value. The following theorem gives similar bounds for Taylor (and hence Maclaurin) polynomials.

THEOREM 76: TAYLOR'S THEOREM

1. Let f be a function whose $n+1^{\text{th}}$ derivative exists on an interval I and let c be in I. Then, for each x in I, there exists z_x between x and c such that

$$f(x)=f(c)+f'(c)(x-c)+\frac{f'(c)}{2!}(x-c)^2+\cdots+\frac{f^{(n)}(c)}{n!}(x-c)^n+R_n(x), \tag{8.7.9}$$

where $R_n(x)=\dfrac{f^{(n+1)}(z_x)}{(n+1)!}(x-c)^{(n+1)}$.

2. $\left|R_n(x)\right| \le \dfrac{\max\left|f^{(n+1)}(z)\right|}{(n+1)!}\left|(x-c)^{(n+1)}\right|$

The first part of Taylor's Theorem states that $f(x)=p_n(x)+R_n(x)$, where $p_n(x)$ is the n^{th} order Taylor polynomial and $R_n(x)$ is the remainder, or error, in the Taylor approximation. The second part gives bounds on how big that error can be. If the $(n+1)^{\text{th}}$ derivative is large, the error may be large; if x is far from c, the error may also be large. However, the $(n+1)!$ term in the denominator tends to ensure that the error gets smaller as n increases.

The following example computes error estimates for the approximations of $\ln 1.5$ and $\ln 2$ made in Example 8.7.2.

Example 8.7.3: Finding error bounds of a Taylor polynomial

Use Theorem 76 to find error bounds when approximating $\ln 1.5$ and $\ln 2$ with $p_6(x)$, the Taylor polynomial of degree 6 of $f(x) = \ln x$ at $x=1$, as calculated in Example 8.7.2.

Solution

1. We start with the approximation of $\ln 1.5$ with $p_6(1.5)$. The theorem references an open interval I that contains both x and c. The smaller the interval we use the better; it will give us a more accurate (and smaller!) approximation of the error. We let $I = (0.9, 1.6)$, as this interval contains both $c=1$ and $x=1.5$.

   The theorem references $\max\left|f^{(n+1)}(z)\right|$. In our situation, this is asking "How big can the 7^{th} derivative of $y=\ln x$ be on the interval $(0.9,1.6)$?" The seventh derivative is $y=-6!/x^7$. The largest value it attains on I is about 1506. Thus we can bound the error as:

$$\begin{aligned}\left|R_6(1.5)\right| &\le \frac{\max\left|f^{(7)}(z)\right|}{7!}\left|(1.5-1)^7\right| \\ &\le \frac{1506}{5040}\cdot\frac{1}{2^7} \\ &\approx 0.0023.\end{aligned}$$

   We computed $p_6(1.5)=0.404688$; using a calculator, we find $\ln 1.5 \approx 0.405465$, so the actual error is about 0.000778, which is less than our bound of 0.0023. This affirms Taylor's Theorem; the theorem states that our approximation would be within about 2 thousandths of the actual value, whereas the approximation was actually closer.
2. We again find an interval I that contains both $c=1$ and $x=2$; we choose $I=(0.9, 2.1)$. The maximum value of the seventh derivative of f on this interval is again about 1506 (as the largest values come near $x=0.9$). Thus

$$\begin{aligned}\left|R_6(2)\right| &\le \frac{\max\left|f^{(7)}(z)\right|}{7!}\left|(2-1)^7\right| \\ &\le \frac{1506}{5040}\cdot 1^7 \\ &\approx 0.30.\end{aligned}$$

This bound is not as nearly as good as before. Using the degree 6 Taylor polynomial at $x=1$ will bring us within 0.3 of the correct answer. As $p_6(2)\approx 0.61667$, our error estimate guarantees that the actual value of $\ln 2$ is somewhere between 0.31667 and 0.91667. These bounds are not particularly useful.

In reality, our approximation was only off by about 0.07. However, we are approximating ostensibly because we do not know the real answer. In order to be assured that we have a good approximation, we would have to resort to using a polynomial of higher degree.

We practice again. This time, we use Taylor's theorem to find n that guarantees our approximation is within a certain amount.

Example $8.7.4$: Finding sufficiently accurate Taylor polynomials

Find n such that the n^{th} Taylor polynomial of $f(x)=\cos x$ at $x=0$ approximates $\cos 2$ to within 0.001 of the actual answer. What is $p_n(2)$?

Solution

Following Taylor's theorem, we need bounds on the size of the derivatives of $f(x)=\cos x$. In the case of this trigonometric function, this is easy. All derivatives of cosine are $\pm\sin x$ or $\pm\cos x$. In all cases, these functions are never greater than 1 in absolute value. We want the error to be less than 0.001. To find the appropriate n, consider the following inequalities:

$$\begin{aligned}\frac{\max\left|f^{(n+1)}(z)\right|}{(n+1)!}\left|(2-0)^{(n+1)}\right| &\le 0.001 \\ \frac{1}{(n+1)!}\cdot 2^{(n+1)} &\le 0.001\end{aligned}$$

We find an n that satisfies this last inequality with trial-and-error. When $n=8$, we have $\frac{2^{8+1}}{(8+1)!}\approx 0.0014$; when $n=9$, we have $\frac{2^{9+1}}{(9+1)!}\approx 0.000282<0.001$. Thus we want to approximate $\cos 2$ with $p_9(2)$.\

We now set out to compute $p_9(x)$. We again need a table of the derivatives of $f(x)=\cos x$ evaluated at $x=0$. A table of these values is given in Figure $8.7.8$.

$$\begin{array}{lcl}
f(x)=\cos x & \Rightarrow & f(0)=1 \\
f'(x)=-\sin x & \Rightarrow & f'(0)=0 \\
f''(x)=-\cos x & \Rightarrow & f''(0)=-1 \\
f'''(x)=\sin x & \Rightarrow & f'''(0)=0 \\
f^{(4)}(x)=\cos x & \Rightarrow & f^{(4)}(0)=1 \\
f^{(5)}(x)=-\sin x & \Rightarrow & f^{(5)}(0)=0 \\
f^{(6)}(x)=-\cos x & \Rightarrow & f^{(6)}(0)=-1 \\
f^{(7)}(x)=\sin x & \Rightarrow & f^{(7)}(0)=0 \\
f^{(8)}(x)=\cos x & \Rightarrow & f^{(8)}(0)=1 \\
f^{(9)}(x)=-\sin x & \Rightarrow & f^{(9)}(0)=0
\end{array}$$

Figure $8.7.9$: A table of the derivatives of $f(x)=cosx$ evaluated at $x=0$.

Notice how the derivatives, evaluated at $x=0$, follow a certain pattern. All the odd powers of x in the Taylor polynomial will disappear as their coefficient is 0. While our error bounds state that we need $p_9(x)$, our work shows that this will be the same as $p_8(x)$.

Since we are forming our polynomial at $x=0$, we are creating a Maclaurin polynomial, and:

CC BY NC

$$\begin{aligned} p_8(x) &= f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \cdots + \frac{f^{(8)}}{8!}x^8 \\ &= 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \frac{1}{6!}x^6 + \frac{1}{8!}x^8 \end{aligned}$$

We finally approximate $\cos 2$:

$$\cos 2 \approx p_8(2) = -\frac{131}{315} \approx -0.41587.$$

Our error bound guarantee that this approximation is within 0.001 of the correct answer. Technology shows us that our approximation is actually within about 0.0003 of the correct answer.

Figure $8.7.10$ shows a graph of $y = p_8(x)$ and $y = \cos x$. Note how well the two functions agree on about $(-\pi, \pi)$.

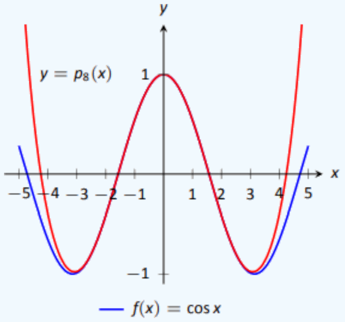


Figure $8.7.10$: A graph of $f(x) = cos x$ and its degree 8 Maclaurin polynomial.

Example $8.7.5$: Finding and using Taylor polynomials

1. Find the degree 4 Taylor polynomial, $p_4(x)$, for $f(x) = \sqrt{x}$ at $x = 4$.
2. Use $p_4(x)$ to approximate $\sqrt{3}$.
3. Find bounds on the error when approximating $\sqrt{3}$ with $p_4(3)$.

Solution

$$\begin{array}{lll} f(x) = \sqrt{x} & \Rightarrow & f(4) = 2 \\ f'(x) = \dfrac{1}{2\sqrt{x}} & \Rightarrow & f'(4) = \dfrac{1}{4} \\ f''(x) = \dfrac{-1}{4x^{3/2}} & \Rightarrow & f''(4) = \dfrac{-1}{32} \\ f'''(x) = \dfrac{3}{8x^{5/2}} & \Rightarrow & f'''(4) = \dfrac{3}{256} \\ f^{(4)}(x) = \dfrac{-15}{16x^{7/2}} & \Rightarrow & f^{(4)}(4) = \dfrac{-15}{2048} \end{array}$$

Figure $8.7.11$: A table of the derivatives of $f(x) = \sqrt{x}$ evaluated at $x = 4$.

1. We begin by evaluating the derivatives of f at $x = 4$. This is done in Figure $8.7.11$. These values allow us to form the Taylor polynomial $p_4(x)$:

$$\begin{aligned} &p_4(x) \\ &= 2 + \frac{1}{4}(x-4) + \frac{-1/32}{2!}(x-4)^2 + \frac{3/256}{3!}(x-4)^3 + \frac{-15/2048}{4!}(x-4)^4. \end{aligned} \tag{8.7.10}$$

2. As $p_4(x) \approx \sqrt{x}$ near $x = 4$, we approximate $\sqrt{3}$ with $p_4(3) = 1.73212$.
3. To find a bound on the error, we need an open interval that contains $x = 3$ and $x = 4$. We set $I = (2.9, 4.1)$. The largest value the fifth derivative of $f(x) = \sqrt{x}$ takes on this interval is near $x = 2.9$, at about 0.0273. Thus

$$\left|R_4(3)\right| \leq \frac{0.0273}{5!}\left|(3-4)^5\right| \approx 0.00023. \tag{8.7.11}$$

This shows our approximation is accurate to at least the first 2 places after the decimal. (It turns out that our approximation is actually accurate to 4 places after the decimal.) A graph of $f(x) = \sqrt{x}$ and $p_4(x)$ is given in Figure $8.7.12$. Note how the two functions are nearly indistinguishable on $(2,7)$.

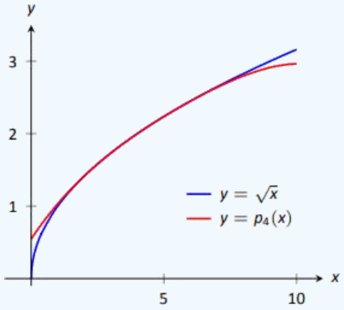


Figure $8.7.12$: A graph of $f(x)=\sqrt{x}$ and its degree 4 Taylor polynomial at $x=4$.

Our final example gives a brief introduction to using Taylor polynomials to solve differential equations.

Example $8.7.6$: Approximating an unknown function

A function $y=f(x)$ is unknown save for the following two facts.

1. $y(0)=f(0)=1$, and
2. $y'=y^2$

(This second fact says that amazingly, the derivative of the function is actually the function squared!)

Find the degree 3 Maclaurin polynomial $p_3(x)$ of $y=f(x)$.

Solution

One might initially think that not enough information is given to find $p_3(x)$. However, note how the second fact above actually lets us know what $y'(0)$ is:

$$y'=y^2 \Rightarrow y'(0)=y^2(0).$$

Since $y(0)=1$, we conclude that $y'(0)=1$.

Now we find information about y''. Starting with $y'=y^2$, take derivatives of both sides, *with respect to* x. That means we must use implicit differentiation.

$$\begin{aligned} y' &= y^2 \\ \frac{d}{dx}(y') &= \frac{d}{dx}(y^2) \\ y'' &= 2y\cdot y'. \\ \text{Now evaluate both sides at } x=0\text{:} & \\ y''(0) &= 2y(0)\cdot y'(0) \\ y''(0) &= 2 \end{aligned}$$

We repeat this once more to find $y'''(0)$. We again use implicit differentiation; this time the Product Rule is also required.

$$\begin{aligned} \frac{d}{dx}(y'') &= \frac{d}{dx}(2yy') \\ y''' &= 2y'\cdot y' + 2y\cdot y''. \\ \text{Now evaluate both sides at } x=0\text{:} & \\ y'''(0) &= 2y'(0)^2 + 2y(0)y''(0) \\ y'''(0) &= 2+4=6 \end{aligned}$$

In summary, we have:

$$y(0)=1 \qquad y'(0)=1 \qquad y''(0)=2 \qquad y'''(0)=6.$$

We can now form $p_3(x)$:

$$\begin{aligned} p_3(x) &= 1+x+\frac{2}{2!}x^2+\frac{6}{3!}x^3 \\ &= 1+x+x^2+x^3. \end{aligned}$$

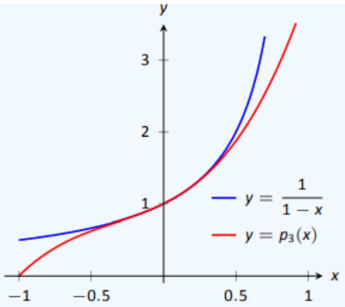


Figure 8.7.13: A graph of $y=-1/(x-1)$ and $y=p_a(x)$ from Example 8.7.6.

It turns out that the differential equation we started with, $y'=y^2$, where $y(0)=1$, can be solved without too much difficulty: $y = \dfrac{1}{1-x}$. Figure 8.28 shows this function plotted with $p_3(x)$. Note how similar they are near $x=0$.

It is beyond the scope of this text to pursue error analysis when using Taylor polynomials to approximate solutions to differential equations. This topic is often broached in introductory Differential Equations courses and usually covered in depth in Numerical Analysis courses. Such an analysis is very important; one needs to know how good their approximation is. We explored this example simply to demonstrate the usefulness of Taylor polynomials.

Most of this chapter has been devoted to the study of infinite series. This section has taken a step back from this study, focusing instead on finite summation of terms. In the next section, we explore **Taylor Series**, where we represent a function with an infinite series.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 8.7: Taylor Polynomials is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

8.8: Taylor Series

In Section 8.6, we showed how certain functions can be represented by a power series function. In 8.7, we showed how we can approximate functions with polynomials, given that enough derivative information is available. In this section we combine these concepts: if a function $f(x)$ is infinitely differentiable, we show how to represent it with a power series function.

Definition 39 taylor and maclaurin series

Let $f(x)$ have derivatives of all orders at $x=c$.

1. The **Taylor Series of** $f(x)$, **centered at** c is

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(c)}{n!}(x-c)^n. \tag{8.8.1}$$

2. Setting $c=0$ gives the **Maclaurin Series of** $f(x)$:

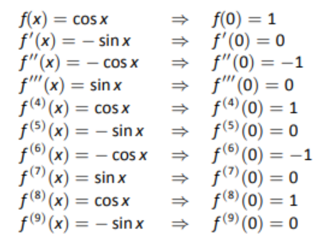
$$\sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!}x^n. \tag{8.8.2}$$

The difference between a *Taylor polynomial* and a *Taylor series* is the former is a polynomial, containing only a finite number of terms, whereas the latter is a series, a summation of an infinite set of terms. When creating the Taylor polynomial of degree n for a function $f(x)$ at $x=c$,we needed to evaluate f,and the first n derivatives of f,at $x=c$.When creating the Taylor series of f, it helps to find a pattern that describes the n^{th} derivative of f at $x=c$.We demonstrate this in the next two examples.

Example 8.8.1: The Maclaurin series of $f(x)=\cos x$

Find the Maclaurin series of $f(x)=\cos x$.

Solution
In Example 8.7.4 we found the 8^{th} degree Maclaurin polynomial of $\cos x$.In doing so, we created the table shown in Figure 8.29.



$f(x)=\cos x \quad \Rightarrow \quad f(0)=1$
$f'(x)=-\sin x \quad \Rightarrow \quad f'(0)=0$
$f''(x)=-\cos x \quad \Rightarrow \quad f''(0)=-1$
$f'''(x)=\sin x \quad \Rightarrow \quad f'''(0)=0$
$f^{(4)}(x)=\cos x \quad \Rightarrow \quad f^{(4)}(0)=1$
$f^{(5)}(x)=-\sin x \quad \Rightarrow \quad f^{(5)}(0)=0$
$f^{(6)}(x)=-\cos x \quad \Rightarrow \quad f^{(6)}(0)=-1$
$f^{(7)}(x)=\sin x \quad \Rightarrow \quad f^{(7)}(0)=0$
$f^{(8)}(x)=\cos x \quad \Rightarrow \quad f^{(8)}(0)=1$
$f^{(9)}(x)=-\sin x \quad \Rightarrow \quad f^{(9)}(0)=0$

Figure 8.29: A table of the derivatives of $f(x)=\cos x$ evaluated at $x=0$.

Notice how $f^{(n)}(0)=0$ when n is odd, $f^{(n)}(0)=1$ when n is divisible by 4,and $f^{(n)}(0)=-1$ when n is even but not divisible by 4. Thus the Maclaurin series of $\cos x$ is

$$1-\frac{x^2}{2}+\frac{x^4}{4!}-\frac{x^6}{6!}+\frac{x^8}{8!}-\cdots \tag{8.8.3}$$

We can go further and write this as a summation. Since we only need the terms where the power of x is even, we write the power series in terms of x^{2n}:

$$\sum_{n=0}^{\infty}(-1)^n\frac{x^{2n}}{(2n)!}. \tag{8.8.4}$$

Example 8.8.2: The Taylor series of $f(x)=\ln x$ at $x=1$

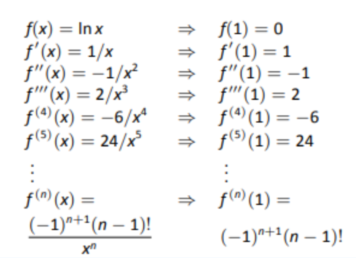
Find the Taylor series of $f(x)=\ln x$ centered at $x=1$.

Solution

Figure 8.30 shows the n^{th} derivative of $\ln x$ evaluated at $x=1$ for $n=0,\ldots,5$,along with an expression for the n^{th} term:

$$f^{(n)}(1)=(-1)^{n+1}(n-1)! \quad \text{for } n\geq 1. \tag{8.8.5}$$

Remember that this is what distinguishes Taylor series from Taylor polynomials; we are very interested in finding a pattern for the n^{th} term, not just finding a finite set of coefficients for a polynomial.



$$\begin{array}{lcl}
f(x)=\ln x & \Rightarrow & f(1)=0 \\
f'(x)=1/x & \Rightarrow & f'(1)=1 \\
f''(x)=-1/x^2 & \Rightarrow & f''(1)=-1 \\
f'''(x)=2/x^3 & \Rightarrow & f'''(1)=2 \\
f^{(4)}(x)=-6/x^4 & \Rightarrow & f^{(4)}(1)=-6 \\
f^{(5)}(x)=24/x^5 & \Rightarrow & f^{(5)}(1)=24 \\
\vdots & & \vdots \\
f^{(n)}(x)= & \Rightarrow & f^{(n)}(1)= \\
\frac{(-1)^{n+1}(n-1)!}{x^n} & & (-1)^{n+1}(n-1)!
\end{array}$$

Figure 8.30: Derivatives of $\ln x$ evaluated at $x=1$.

Since $f(1)=\ln 1=0$,we skip the first term and start the summation with $n=1$,giving the Taylor series for $\ln x$,centered at $x=1$,as

$$\sum_{n=1}^{\infty}(-1)^{n+1}(n-1)!\frac{1}{n!}(x-1)^n=\sum_{n=1}^{\infty}(-1)^{n+1}\frac{(x-1)^n}{n}. \tag{8.8.6}$$

It is important to note that Definition 39 defines a Taylor series given a function $f(x)$; however, we *cannot* yet state that $f(x)$ *is equal* to its Taylor series. We will find that "most of the time" they are equal, but we need to consider the conditions that allow us to conclude this.

Theorem 76 states that the error between a function $f(x)$ and its n^{th}--degree Taylor polynomial $p_n(x)$ is $R_n(x)$,where

$$|R_n(x)|\leq\frac{\max\left|f^{(n+1)}(z)\right|}{(n+1)!}\left|(x-c)^{(n+1)}\right|. \tag{8.8.7}$$

If $R_n(x)$ goes to 0 for each x in an interval I as n approaches infinity, we conclude that the function is equal to its Taylor series expansion.

theorem 77 function and taylor series equality

Let $f(x)$ have derivatives of all orders at $x=c$,let $R_n(x)$ be as stated in Theorem 76, and let I be an interval on which the Taylor series of $f(x)$ converges.

If $\lim\limits_{n\to\infty}R_n(x)=0$ for all x in I,then

$$f(x)=\sum_{n=0}^{\infty}\frac{f^{(n)}(c)}{n!}(x-c)^n \quad \text{on } I. \tag{8.8.8}$$

We demonstrate the use of this theorem in an example.

Example 8.8.3: Establishing equality of a function and its Taylor series

Show that $f(x)=\cos x$ is equal to its Maclaurin series, as found in Example 8.8.1, for all x.

Solution

Given a value x,the magnitude of the error term $R_n(x)$ is bounded by

$$|R_n(x)|\leq\frac{\max\left|f^{(n+1)}(z)\right|}{(n+1)!}\left|x^{n+1}\right|. \tag{8.8.9}$$

Since all derivatives of $\cos x$ are $\pm\sin x$ or $\pm\cos x$,whose magnitudes are bounded by 1,we can state

$$|R_n(x)|\leq\frac{1}{\quad}\left|x^{n+1}\right| \tag{8.8.10}$$

$$|R_n(x)| \leq \frac{}{(n+1)!}|x^{n+1}| \tag{8.8.10}$$

which implies

$$-\frac{|x^{n+1}|}{(n+1)!} \leq R_n(x) \leq \frac{|x^{n+1}|}{(n+1)!}. \tag{8.8.11}$$

For any x, $\lim\limits_{n\to\infty} \frac{x^{n+1}}{(n+1)!} = 0$. Applying the Squeeze Theorem to Equation 8.8.11, we conclude that $\lim\limits_{n\to\infty} R_n(x) = 0$ for all x,and hence

$$\cos x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!} \quad \text{for all } x. \tag{8.8.12}$$

It is natural to assume that a function is equal to its Taylor series on the series' interval of convergence, but this is not the case. In order to properly establish equality, one must use Theorem 77. This is a bit disappointing, as we developed beautiful techniques for determining the interval of convergence of a power series, and proving that $R_n(x) \to 0$ can be cumbersome as it deals with high order derivatives of the function.

There is good news. A function $f(x)$ that is equal to its Taylor series, centered at any point the domain of $f(x)$,is said to be an **analytic function**, and most, if not all, functions that we encounter within this course are analytic functions. Generally speaking, any function that one creates with elementary functions (polynomials, exponentials, trigonometric functions, etc.) that is not piecewise defined is probably analytic. For most functions, we assume the function is equal to its Taylor series on the series' interval of convergence and only use Theorem 77 when we suspect something may not work as expected.

We develop the Taylor series for one more important function, then give a table of the Taylor series for a number of common functions.

Example $8.8.4$: The Binomial Series

Find the Maclaurin series of $f(x) = (1+x)^k$,$k \neq 0$.

Solution

When k is a positive integer, the Maclaurin series is finite. For instance, when $k=4$,we have

$$f(x) = (1+x)^4 = 1 + 4x + 6x^2 + 4x^3 + x^4. \tag{8.8.13}$$

The coefficients of x when k is a positive integer are known as the *binomial coefficients*, giving the series we are developing its name.

When $k = 1/2$,we have $f(x) = \sqrt{1+x}$.Knowing a series representation of this function would give a useful way of approximating $\sqrt{1.3}$,for instance.

To develop the Maclaurin series for $f(x) = (1+x)^k$ for any value of $k \neq 0$,we consider the derivatives of f evaluated at $x=0$:

$$\begin{aligned}
f(x) &= (1+x)^k & f(0) &= 1 \\
f'(x) &= k(1+x)^{k-1} & f'(0) &= k \\
f''(x) &= k(k-1)(1+x)^{k-2} & f''(0) &= k(k-1) \\
f'''(x) &= k(k-1)(k-2)(1+x)^{k-3} & f'''(0) &= k(k-1)(k-2) \\
&\vdots & &\vdots \\
f^{(n)}(x) &= k(k-1)\cdots\big(k-(n-1)\big)(1+x)^{k-n} & f^{(n)}(0) &= k(k-1)\cdots\big(k-(n-1)\big)
\end{aligned}$$

Thus the Maclaurin series for $f(x) = (1+x)^k$ is

$$1 + k + \frac{k(k-1)}{2!} + \frac{k(k-1)(k-2)}{3!} + \ldots + \frac{k(k-1)\cdots\big(k-(n-1)\big)}{n!} + \ldots \tag{8.8.14}$$

It is important to determine the interval of convergence of this series. With

$$a_n = \frac{k(k-1)\cdots\big(k-(n-1)\big)}{}x^n \tag{8.8.15}$$

we apply the Ratio Test:

$$\begin{aligned}\lim_{n\to\infty}\frac{|a_{n+1}|}{|a_n|} &= \lim_{n\to\infty}\left|\frac{k(k-1)\cdots(k-n)}{(n+1)!}x^{n+1}\right| \bigg/ \left|\frac{k(k-1)\cdots\big(k-(n-1)\big)}{n!}x^n\right| \\ &= \lim_{n\to\infty}\left|\frac{k-n}{n}x\right| \\ &= |x|.\end{aligned}$$

The series converges absolutely when the limit of the Ratio Test is less than 1; therefore, we have absolute convergence when $|x| < 1$.

While outside the scope of this text, the interval of convergence depends on the value of k.When $k > 0$,the interval of convergence is $[-1,1]$.When $-1 < k < 0$,the interval of convergence is $[-1,1)$.If $k \leq -1$,the interval of convergence is $(-1,1)$.

We learned that Taylor polynomials offer a way of approximating a "difficult to compute" function with a polynomial. Taylor series offer a way of exactly representing a function with a series. One probably can see the use of a good approximation; is there any use of representing a function exactly as a series?

While we should not overlook the mathematical beauty of Taylor series (which is reason enough to study them), there are practical uses as well. They provide a valuable tool for solving a variety of problems, including problems relating to integration and differential equations.

In Key Idea 32 (on the following page) we give a table of the Taylor series of a number of common functions. We then give a theorem about the "algebra of power series," that is, how we can combine power series to create power series of new functions. This allows us to find the Taylor series of functions like $f(x) = e^x \cos x$ by knowing the Taylor series of e^x and $\cos x$.

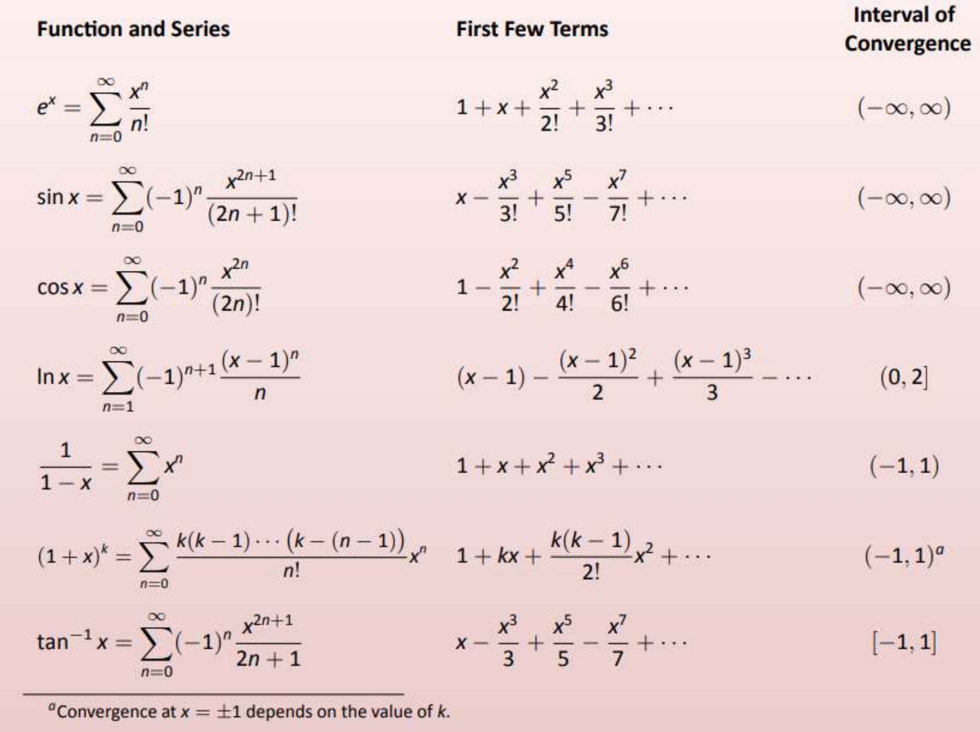
Before we investigate combining functions, consider the Taylor series for the arctangent function (see Key Idea 32). Knowing that $\tan^{-1}(1) = \pi/4$,we can use this series to approximate the value of π:

$$\frac{\pi}{4} = \tan^{-1}(1) = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \cdots \tag{8.8.16}$$

$$\pi = 4\left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \cdots\right) \tag{8.8.17}$$

Unfortunately, this particular expansion of π converges very slowly. The first 100 terms approximate π as 3.13159,which is not particularly good.

KEY IDEA 32 IMPORTANT TAYLOR SERIES EXPASIONS



Function and Series	First Few Terms	Interval of Convergence
$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$	$1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$	$(-\infty, \infty)$
$\sin x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}$	$x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$	$(-\infty, \infty)$
$\cos x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}$	$1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots$	$(-\infty, \infty)$
$\ln x = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{(x-1)^n}{n}$	$(x-1) - \frac{(x-1)^2}{2} + \frac{(x-1)^3}{3} - \cdots$	$(0, 2]$
$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$	$1 + x + x^2 + x^3 + \cdots$	$(-1, 1)$
$(1+x)^k = \sum_{n=0}^{\infty} \frac{k(k-1)\cdots\left(k-(n-1)\right)}{n!} x^n$	$1 + kx + \frac{k(k-1)}{2!} x^2 + \cdots$	$(-1, 1)^a$
$\tan^{-1} x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1}$	$x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots$	$[-1, 1]$

---

[a] Convergence at $x = \pm 1$ depends on the value of k.

THEOREM 78 ALGEBRA OF POWER SERIES

Let $f(x) = \sum_{n=0}^{\infty} a_n x^n$ and $g(x) = \sum_{n=0}^{\infty} b_n x^n$ converge absolutely for $|x| < R$,and let $h(x)$ be continuous.

1. $f(x) \pm g(x) = \sum_{n=0}^{\infty} (a_n \pm b_n) x^n$ \quad for $|x| < R$.
2. $f(x)g(x) = \left(\sum_{n=0}^{\infty} a_n x^n\right)\left(\sum_{n=0}^{\infty} b_n x^n\right) = \sum_{n=0}^{\infty} \left(a_0 b_n + a_1 b_{n-1} + \ldots a_n b_0\right) x^n$ for $|x| < R$.
3. $f\left(h(x)\right) = \sum_{n=0}^{\infty} a_n \left(h(x)\right)^n \quad$ for $|h(x)| < R$.

Example 8.8.5: Combining Taylor series

Write out the first 3 terms of the Taylor Series for $f(x) = e^x \cos x$ using Key Idea 32 and Theorem 78.

Solution

Key Idea 32 informs us that

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \quad \text{and} \quad \cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots. \tag{8.8.18}$$

Applying Theorem 78, we find that

$$e^x \cos x = \left(1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots\right)\left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots\right). \tag{8.8.19}$$

$$\text{Distribute the right hand expression across the left:} \tag{8.8.20}$$

$$= 1\left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots\right) + x\left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots\right) + \frac{x^2}{2!}\left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots\right) \tag{8.8.21}$$

$$+ \frac{x^3}{3!}\left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots\right) + \frac{x^4}{4!}\left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots\right) + \cdots \tag{8.8.22}$$

$$\text{Distribute again and collect like terms.} \tag{8.8.23}$$

$$= 1 + x - \frac{x^3}{3} - \frac{x^4}{6} - \frac{x^5}{30} + \frac{x^7}{630} + \cdots \tag{8.8.24}$$

While this process is a bit tedious, it is much faster than evaluating all the necessary derivatives of $e^x \cos x$ and computing the Taylor series directly.

Because the series for e^x and $\cos x$ both converge on $(-\infty, \infty)$,so does the series expansion for $e^x \cos x$.

Example 8.8.6: Creating new Taylor series

Use Theorem 78 to create series for $y=\sin(x^2)$ and $y=\ln(\sqrt{x})$.

Solution
Given that

$$\sin x=\sum_{n=0}^{\infty}(-1)^n\frac{x^{2n+1}}{(2n+1)!}=x-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\cdots, \tag{8.8.25}$$

we simply substitute x^2 for x in the series, giving

$$\sin(x^2)=\sum_{n=0}^{\infty}(-1)^n\frac{(x^2)^{2n+1}}{(2n+1)!}=x^2-\frac{x^6}{3!}+\frac{x^{10}}{5!}-\frac{x^{14}}{7!}\cdots. \tag{8.8.26}$$

Since the Taylor series for $\sin x$ has an infinite radius of convergence, so does the Taylor series for $\sin(x^2)$.

The Taylor expansion for $\ln x$ given in Key Idea 32 is centered at $x=1$,so we will center the series for $\ln(\sqrt{x})$ at $x=1$ as well. With

$$\ln x=\sum_{n=1}^{\infty}(-1)^{n+1}\frac{(x-1)^n}{n}=(x-1)-\frac{(x-1)^2}{2}+\frac{(x-1)^3}{3}-\cdots, \tag{8.8.27}$$

we substitute $\sqrt{x}$ for x to obtain

$$\ln(\sqrt{x})=\sum_{n=1}^{\infty}(-1)^{n+1}\frac{(\sqrt{x}-1)^n}{n}=(\sqrt{x}-1)-\frac{(\sqrt{x}-1)^2}{2}+\frac{(\sqrt{x}-1)^3}{3}-\cdots. \tag{8.8.28}$$

While this is not strictly a power series, it is a series that allows us to study the function $\ln(\sqrt{x})$.Since the interval of convergence of $\ln x$ is $(0,2]$,and the range of $\sqrt{x}$ on $(0,4]$ is $(0,2]$,the interval of convergence of this series expansion of $\ln(\sqrt{x})$ is $(0,4]$.

Note: In Example 8.8.6, one could create a series for $\ln(\sqrt{x})$ by simply recognizing that $\ln(\sqrt{x})=\ln(x^{1/2})=1/2\ln x$,and hence multiplying the Taylor series for $\ln x$ by $1/2$.This example was chosen to demonstrate other aspects of series, such as the fact that the interval of convergence changes.

Example 8.8.7: Using Taylor series to evaluate definite integrals

Use the Taylor series of e^{-x^2} to evaluate $\int_0^1 e^{-x^2}\,dx$.

Solution

We learned, when studying Numerical Integration, that e^{-x^2} does not have an antiderivative expressible in terms of elementary functions. This means any definite integral of this function must have its value approximated, and not computed exactly.

We can quickly write out the Taylor series for e^{-x^2} using the Taylor series of e^x:

$$e^x=\sum_{n=0}^{\infty}\frac{x^n}{n!}=1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\cdots$$

and so

$$\begin{aligned}e^{-x^2}&=\sum_{n=0}^{\infty}\frac{(-x^2)^n}{n!}\\&=\sum_{n=0}^{\infty}(-1)^n\frac{x^{2n}}{n!}\\&=1-x^2+\frac{x^4}{2!}-\frac{x^6}{3!}+\cdots.\end{aligned}$$

We use Theorem 75 to integrate:

$$\int e^{-x^2}\,dx=C+x-\frac{x^3}{3}+\frac{x^5}{5\cdot 2!}-\frac{x^7}{7\cdot 3!}+\cdots+(-1)^n\frac{x^{2n+1}}{(2n+1)n!}+\cdots \tag{8.8.29}$$

This *is* the antiderivative of e^{-x^2};while we can write it out as a series, we cannot write it out in terms of elementary functions. We can evaluate the definite integral $\int_0^1 e^{-x^2}\,dx$ using this antiderivative; substituting 1 and 0 for x and subtracting gives

$$\int_0^1 e^{-x^2}\,dx = 1-\frac{1}{3}+\frac{1}{5\cdot 2!}-\frac{1}{7\cdot 3!}+\frac{1}{9\cdot 4!}\cdots. \tag{8.8.30}$$

Summing the 5 terms shown above give the approximation of 0.74749. Since this is an alternating series, we can use the Alternating Series Approximation Theorem, (Theorem 71), to determine how accurate this approximation is. The next term of the series is $1/(11\cdot 5!)\approx 0.00075758$.Thus we know our approximation is within 0.00075758 of the actual value of the integral. This is arguably much less work than using Simpson's Rule to approximate the value of the integral.

Example $8.8.8$: Using Taylor series to solve differential equations

Solve the differential equation $y' = 2y$ in terms of a power series, and use the theory of Taylor series to recognize the solution in terms of an elementary function.

Solution

We found the first 5 terms of the power series solution to this differential equation in Example 8.6.5 in Section 8.6. These are:

$$a_0=1,\quad a_1=2,\quad a_2=\frac{4}{2}=2,\quad a_3=\frac{8}{2\cdot 3}=\frac{4}{3},\quad a_4=\frac{16}{2\cdot 3\cdot 4}=\frac{2}{3}. \tag{8.8.31}$$

We include the "unsimplified" expressions for the coefficients found in Example 8.6.5 as we are looking for a pattern. It can be shown that $a_n = 2^n/n!$.Thus the solution, written as a power series, is

$$y=\sum_{n=0}^{\infty}\frac{2^n}{n!}x^n=\sum_{n=0}^{\infty}\frac{(2x)^n}{n!}. \tag{8.8.32}$$

Using Key Idea 32 and Theorem 78, we recognize $f(x) = e^{2x}$:

$$e^x=\sum_{n=0}^{\infty}\frac{x^n}{n!}\qquad\Rightarrow\qquad e^{2x}=\sum_{n=0}^{\infty}\frac{(2x)^n}{n!}. \tag{8.8.33}$$

Finding a pattern in the coefficients that match the series expansion of a known function, such as those shown in Key Idea 32, can be difficult. What if the coefficients in the previous example were given in their reduced form; how could we still recover the function $y = e^{2x}$?

Suppose that all we know is that

$$a_0=1,\quad a_1=2,\quad a_2=2,\quad a_3=\frac{4}{3},\quad a_4=\frac{2}{3}. \tag{8.8.34}$$

Definition 39 states that each term of the Taylor expansion of a function includes an $n!$.This allows us to say that

$$a_2=2=\frac{b_2}{2!},\quad a_3=\frac{4}{3}=\frac{b_3}{3!},\quad\text{and}\quad a_4=\frac{2}{3}=\frac{b_4}{4!} \tag{8.8.35}$$

for some values b_2,b_3 and b_4.

Solving for these values, we see that $b_2 = 4$,$b_3 = 8$ and $b_4 = 16$.That is, we are recovering the pattern we had previously seen, allowing us to write

$$\begin{aligned} f(x) &= \sum_{n=0}^{\infty}a_nx^n=\sum_{n=0}^{\infty}\frac{b_n}{n!}x^n \\ &= 1+2x+\frac{4}{2!}x^2+\frac{8}{3!}x^3+\frac{16}{4!}x^4+\cdots \end{aligned}$$

From here it is easier to recognize that the series is describing an exponential function.

There are simpler, more direct ways of solving the differential equation $y' = 2y$.We applied power series techniques to this equation to demonstrate its utility, and went on to show how *sometimes* we are able to recover the solution in terms of elementary functions using the theory of Taylor series. Most differential equations faced in real scientific and engineering situations are much more complicated than this one, but power series can offer a valuable tool in finding, or at least approximating, the solution.

This chapter introduced sequences, which are ordered lists of numbers, followed by series, wherein we add up the terms of a sequence. We quickly saw that such sums do not always add up to "infinity," but rather converge. We studied tests for convergence, then ended the chapter with a formal way of defining functions based on series. Such "series--defined functions" are a valuable tool in solving a number of different problems throughout science and engineering.

Coming in the next chapters are new ways of defining curves in the plane apart from using functions of the form $y=f(x)$.Curves created by these new methods can be beautiful, useful, and important.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 8.8: Taylor Series is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

8.E: Applications of Sequences and Series (Exercises)

8.1: Sequences

Terms and Concepts

1. Use your own words to define a sequence.

2. The domain of a sequence is the ______ numbers.

3. Use your own words to describe the range of a sequence.

4. Describe what it means for a sequence to be bounded.

Problems

In Exercises 5-8, give the first five terms of the given sequence.

5. $a_n = \left\{\frac{4^n}{(n+1)!}\right\}$

6. $b_n = \{(-\frac{3}{2})^n\}$

7. $c_n = \left\{-\frac{n^{n+1}}{n+2}\right\}$

8. $d_n = \left\{\frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^2\right)\right\}$

In Exercises 9-12, determine the n^{th} term of the given sequence.

9. 4, 7, 10, 13, 16, ...

10. $3, -\frac{3}{2}, \frac{3}{4}, -\frac{3}{8}, \ldots$

11. 10, 20, 40, 80, 160, ...

12. $1, 1, \frac{1}{2}, \frac{1}{6}, \frac{1}{24}, \frac{1}{120}, \ldots$

In Exercises 13-16, use the following information to determine the limit of the given sequences.

- $a_n = \left\{\frac{2^n-20}{2^n}\right\}; \quad \lim\limits_{n\to\infty} a_n = 1$
- $b_n = \left\{\left(1+\frac{2}{n}\right)^n\right\}; \quad \lim\limits_{n\to\infty} b_n = e^2$
- $c_n = \left\{\frac{2^n-20}{2^n}\right\}; \quad \lim\limits_{n\to\infty} c_n = 0$

13. $a_n = \left\{\frac{2^n-20}{7\cdot 2^n}\right\}$

14. $a_n = 3b_n - a_n$

15. $a_n = \{\sum(3/n)\left(1 = \frac{2}{n}\right)^n\}$

16. $a_n = \left\{\left(1 = \frac{2}{n}\right)^{2n}\right\}$

In Exercises 17-28, determine whether the sequence converges or diverges. If convergent, give the limit of the sequence.

17. $a_n = \left\{(-1)^n \frac{n}{n+1}\right\}$

18. $a_n = \left\{\frac{4n^2-n+5}{3n^2+1}\right\}$

19. $a_n = \left\{\frac{4^n}{5^n}\right\}$

20. $a_n = \left\{\frac{n-1}{n} - \frac{n}{n-1}\right\}, n \geq 2$

21. $a_n = \{\ln(n)\}$

22. $a_n = \left\{\frac{3n}{\sqrt{n^2+1}}\right\}$

23. $a_n = \left\{\left(1+\frac{1}{n}\right)^n\right\}$

24. $a_n = \left\{5 - \frac{1}{n}\right\}$

25. $a_n = \left\{\frac{(-1)^{n+1}}{n}\right\}$

26. $a_n = \left\{\frac{1.1^n}{n}\right\}$

27. $a_n = \left\{\frac{2n}{n+1}\right\}$

28. $a_n = \left\{(-1)^n \frac{n^2}{2^n-1}\right\}$

In Exercises 29-34, determine whether the sequence is bounded, bounded above, bounded below, or none of the above.

29. $a_n = \sum n$

30. $a_n = \tan n$

31. $a_n = (-1)^n \frac{3n-1}{n}$

32. $a_n = \left\{\frac{3n^2-1}{n}\right\}$

33. $a_n = n \cos n$

34. $a_n = 2^n - n!$

In Exercises 35-38, determine whether the sequence is monotonically increasing or decreasing. If it is not, determine if there is an *m* such that it is monotonic for all $n \geq m$.

35. $a_n = \left\{\frac{n}{n+2}\right\}$

36. $a_n = \left\{\frac{n^2-6n+9}{n}\right\}$

37. $a_n = \left\{(-1)^n \frac{1}{n^3}\right\}$

38. $a_n = \left\{\frac{n^2}{2^n}\right\}$

39. Prove Theorem 56; that is, use the definition of the limit of a sequence to show that if $\lim\limits_{n\to\infty} |a_n| = 0$, then $\lim\limits_{n\to\infty} a_n = 0$.

40. Let $a_n \text{ and } b_n$ be sequences such that $\lim\limits_{n\to\infty} a_n = L \text{ and } \lim\limits_{n\to\infty} b_n = K$.
(a) Show that if $a_n < b_n$ for all *n*, then $L \leq K$.
(b) Give an example where $L = K$.

41. Prove the Squeeze Theorem for sequences: Let $a_n \text{ and } b_n$ be such that $\lim\limits_{n\to\infty} a_n = L \text{ and } \lim\limits_{n\to\infty} b_n = L$, and let c_n be such that $a_n \leq c_n \leq b_n$ for all *n*. Then $\lim\limits_{n\to\infty} c_n = L$

8.2: Infinite Series

Terms and Concepts

1. Use your own words to describe how sequences and series are related.

2. Use your own words to define a partial sum.

3. Given a series $\sum\limits_{n=1}^{\infty} a_n$m describe the two sequences related to the series that are important.

4. Use your own words to explain what a geometric series is.

5. T/F: If a_n is convergent, then $\sum\limits_{n=1}^{\infty} a_n$ is also convergent.

Problems

In Exercises 6-13, a series $\sum\limits_{n=1}^{\infty} a_n$ **is given.**

(a) Give the first 5 partial sums of the series.
(b) Give a graph of the first 5 terms of $a_n \text{ and } S_n$ **on the same axes.**

6. $\sum\limits_{n=1}^{\infty} \frac{(-1)^n}{n}$

7. $\sum\limits_{n=1}^{\infty} \frac{1}{n^2}$

8. $\sum\limits_{n=1}^{\infty} \cos(\pi n)$

9. $\sum_{n=1}^{\infty} n$

10. $\sum_{n=1}^{\infty} \frac{1}{n!}$

11. $\sum_{n=1}^{\infty} \frac{1}{3^n}$

12. $\sum_{n=1}^{\infty} \left(-\frac{9}{10}\right)^n$

13. $\sum_{n=1}^{\infty} \left(\frac{1}{10}\right)^n$

In Exercises 14-19, use Theorem 63 to show the given series diverges.

14. $\sum_{n=1}^{\infty} \frac{3n^2}{n(n+2)}$

15. $\sum_{n=1}^{\infty} \frac{2^n}{n^2}$

16. $\sum_{n=1}^{\infty} \frac{n!}{10^n}$

17. $\sum_{n=1}^{\infty} \frac{5^n-n^5}{5^n+n^5}$

18. $\sum_{n=1}^{\infty} \frac{2^n+1}{2^{n+1}}$

19. $\sum_{n=1}^{\infty} \left(1+\frac{1}{n}\right)^n$

In Exercises 20-29, state whether the given series converges or diverges.

20. $\sum_{n=1}^{\infty} \frac{1}{n^5}$

21. $\sum_{n=0}^{\infty} \frac{1}{5^n}$

22. $\sum_{n=0}^{\infty} \frac{6^n}{5^n}$

23. $\sum_{n=1}^{\infty} n^{-4}$

24. $\sum_{n=1}^{\infty} \sqrt{n}$

25. $\sum_{n=1}^{\infty} \frac{10}{n!}$

26. $\sum_{n=1}^{\infty} \left(\frac{1}{n!}+\frac{1}{n}\right)$

27. $\sum_{n=1}^{\infty} \frac{2}{(2x+8)^2}$

28. $\sum_{n=1}^{\infty} \frac{1}{2n}$

29. $\sum_{n=1}^{\infty} \frac{1}{2n-1}$

In Exercises 30-44, a series is given.
(a) Find a formula for S_n, the n^{th} partial sum of the series.
(b) Determine whether the series converges or diverges. If it converges, state what it converges to.

30. $\sum_{n=0}^{\infty} \frac{1}{4^n}$

31. $1^3+2^3+3^3+4^3+\ldots$

32. $\sum_{n=1}^{\infty} (-1)^n n\backslash)$

33. $\sum\limits_{n=0}^{\infty} \frac{5}{2^n}$

34. $\sum\limits_{n=1}^{\infty} e^{-n}$

35. $1-\frac{1}{3}+\frac{1}{9}-\frac{1}{27}+\frac{1}{81}+\ldots$

36. $\sum\limits_{n=1}^{\infty} \frac{1}{n(n+1)}$

37. $\sum\limits_{n=1}^{\infty} \frac{3}{n(n+2)}$

38. $\sum\limits_{n=1}^{\infty} \frac{1}{(2x-1)(2x+1)}$

39. $\sum\limits_{n=1}^{\infty} \ln\left(\frac{n}{n+1}\right)$

40. $\sum\limits_{n=1}^{\infty} \frac{2n+1}{n^2(n+1)^2}$

41. $\frac{1}{1\cdot 4}+\frac{1}{2\cdot 5}+\frac{1}{3\cdot 6}+\frac{1}{4\cdot 7}+\ldots$

42. $2+\left(\frac{1}{2}+\frac{1}{3}\right)+\left(\frac{1}{4}+\frac{1}{9}\right)+\left(\frac{1}{8}+\frac{1}{27}+\ldots\right)$

43. $\sum\limits_{n=2}^{\infty} \frac{1}{n^2-1}$

44. $\sum\limits_{n=0}^{\infty} (\sin 1)^n$

45. Break the Harmonic Series into the sum of the odd and even terms:
$\sum\limits_{n=1}^{\infty} \frac{1}{n} = \sum\limits_{n=1}^{\infty} \frac{1}{2n-1} + \sum\limits_{n=1}^{\infty} \frac{1}{2n}$.
The goal is to show that each of the series on the right diverge.
(a) Show why $\sum\limits_{n=1}^{\infty} \frac{1}{2n-1} > \sum\limits_{n=1}^{\infty} \frac{1}{2n}$. (Compare each n^{th} partial sum.)
(b) Show why $\sum\limits_{n=1}^{\infty} \frac{1}{2n-1} < 1 + \sum\limits_{n=1}^{\infty} \frac{1}{2n}$
(c) Explain why (a) and (b) demonstrate that the series of odd terms is convergent, if, and only if, the series of even terms is also convergent. (That is, show both converge or both diverge.)
(d) Explain why knowing the Harmonic Series is divergent determines that the even and odd series are also divergent.

46. Show the series $\sum\limits_{n=1}^{\infty} \frac{n}{(2n-1)(2n+1)}$ diverges.

8.3: Integral and Comparison Tests

Terms and Concepts

1. In order to apply the Integral Test to a sequence A_n, the function $a(n)=a_n$ must be _____, _____ and _____.

2.T/F: The Integral Test can be used to determine the sum of a convergent series.

3.What test(s) in this section do not work well with factorials?

4. Suppose $\sum\limits_{n=0}^{\infty} a_n$ is convergent, and there are sequences b_n and c_n such that $b_n \le a_n \le c_n$ for all *n*. What can be said about the series $\sum\limits_{n=0}^{\infty} b_n$ and $\sum\limits_{n=0}^{\infty} c_n$?

Problems

In Exercises 5-12, use the Integral Test to determine the convergence of the given series.

5. $\sum\limits_{n=1}^{\infty} \frac{1}{2^n}$

6. $\sum\limits_{n=1}^{\infty} \frac{1}{n^4}$

7. $\sum_{n=1}^{\infty} \frac{n}{n^2+1}$

8. $\sum_{n=2}^{\infty} \frac{1}{n \ln n}$

9. $\sum_{n=1}^{\infty} \frac{1}{n^2+1}$

10. $\sum_{n=2}^{\infty} \frac{1}{n(\ln n)^2}$

11. $\sum_{n=1}^{\infty} \frac{n}{2^n}$

12. $\sum_{n=1}^{\infty} \frac{\ln n}{n^3}$

In Exercises 13-22, use the Direct Comparison Test to determine the convergence of the given series; state what series is used for comparison.

13. $\sum_{n=1}^{\infty} \frac{1}{n^2+3n-5}$

14. $\sum_{n=1}^{\infty} \frac{1}{4^n+n^2-n}$

15. $\sum_{n=1}^{\infty} \frac{\ln n}{n}$

16. $\sum_{n=1}^{\infty} \frac{1}{n!+n}$

17. $\sum_{n=2}^{\infty} \frac{1}{\sqrt{n^2-1}}$

18. $\sum_{n=1}^{\infty} \frac{1}{\sqrt{n}-2}$

19. $\sum_{n=1}^{\infty} \frac{n^2+n+1}{2^n}$

20. $\sum_{n=1}^{\infty} \frac{2^n}{5^n+10}$

21. $\sum_{n=2}^{\infty} \frac{n}{n^2-1}$

22. $\sum_{n=2}^{\infty} \frac{1}{n^2 \ln n}$

In Exercises 23-32, use the Limit Comparison Test to determine the convergence of the given series; state what series is used for comparison.

23. $\sum_{n=1}^{\infty} \frac{1}{n^2-3n+5}$

24. $\sum_{n=1}^{\infty} \frac{1}{4^n-n^2}$

25. $\sum_{n=4}^{\infty} \frac{\ln n}{n-3}$

26. $\sum_{n=1}^{\infty} \frac{1}{\sqrt{n^2+n}}$

27. $\sum_{n=1}^{\infty} \frac{1}{n+\sqrt{n}}$

28. $\sum_{n=1}^{\infty} \frac{n-10}{n^2+10n+10}$

29. $\sum_{n=1}^{\infty} \sin(1/n)$

30. $\sum_{n=1}^{\infty} \frac{n+5}{n^3-5}$

31. $\sum_{n=1}^{\infty} \frac{\sqrt{n}+3}{n^2+17}$

32. $\sum_{n=1}^{\infty} \frac{1}{\sqrt{n}+100}$

In Exercises 33-40, determine the convergence of the given series. State the test used; more than one test may be appropriate.

33. $\sum_{n=1}^{\infty} \frac{n^2}{2^n}$

34. $\sum_{n=1}^{\infty} \frac{1}{(2n+5)^3}$

35. $\sum_{n=1}^{\infty} \frac{n!}{10^n}$

36. $\sum_{n=1}^{\infty} \frac{\ln n}{n!}$

37. $\sum_{n=1}^{\infty} \frac{1}{3^n+n}$

38. $\sum_{n=1}^{\infty} \frac{n-2}{10n+5}$

39. $\sum_{n=1}^{\infty} \frac{3^n}{n^3}$

40. $\sum_{n=1}^{\infty} \frac{\cos(1/n)}{\sqrt{n}}$

41. Given that $\sum_{n=1}^{\infty} a_n$ converges, state which of the following series converges, may converge, or does not converge.

(a) $\sum_{n=1}^{\infty} \frac{a_n}{n}$

(b) $\sum_{n=1}^{\infty} a_n a_{n+1}$

(c) $\sum_{n=1}^{\infty} (a_n)^2$

(d) $\sum_{n=1}^{\infty} n a_n$

(e) $\sum_{n=1}^{\infty} \frac{1}{a_n}$

8.4: Ratio and Root Tests

Terms and Concepts

1. The Ratio Test is not effective when the terms of a sequence only contain ______ functions

2. The Ratio Test is most effective when the terms of a sequence contains ____ and/or _____ functions.

3. What three convergence tests do not work well with terms containing factorials?

4. The Root Test works particularly well on series where each term is _____ to a _____.

Problems

In Exercises 5-14, determine the convergence of the given series using the Ratio Test. If the Ratio Test is inconclusive, state so and determine the convergence with another test.

5. $\sum_{n=0}^{\infty} \frac{2n}{n!}$

6. $\sum_{n=0}^{\infty} \frac{5^n-3n}{4^n}$

7. $\sum_{n=0}^{\infty} \frac{n!10^n}{(2n)!}$

8. $\sum_{n=1}^{\infty} \frac{5^n+n^4}{7^n+n^2}$

9. $\sum_{n=1}^{\infty} \frac{1}{n}$

10. $\sum_{n=1}^{\infty} \frac{1}{3n^3+7}$

11. $\sum_{n=1}^{\infty} \frac{10 \cdot 5^n}{7^n - 3}$

12. $\sum_{n=1}^{\infty} n \cdot \left(\frac{3}{5}\right)^n$

13. $\sum_{n=1}^{\infty} \frac{2 \cdot 4 \cdot 6 \cdot 8 \cdots 2n}{3 \cdot 6 \cdot 9 \cdot 12 \cdots 3n}$

14. $\sum_{n=1}^{\infty} \frac{n!}{5 \cdot 10 \cdot 15 \cdots (5n)}$

In Exercises 15-24, determine the convergence of the given series using the Root Test. If the Root Test is inconclusive, state so and determine convergence with another test.

15. $\sum_{n=1}^{\infty} \left(\frac{2n+5}{3n+11}\right)^n$

16. $\sum_{n=1}^{\infty} \left(\frac{0.9n^2-n-3}{n^2+n+3}\right)^n$

17. $\sum_{n=1}^{\infty} \frac{2^n n^2}{3^n}$

18. $\sum_{n=1}^{\infty} \frac{1}{n^n}$

19. $\sum_{n=1}^{\infty} \frac{3^n}{n^2 2^{n+1}}$

20. $\sum_{n=1}^{\infty} \frac{4^{n+7}}{7^n}$

21. $\sum_{n=1}^{\infty} \left(\frac{n^2-n}{n^2+n}\right)$

22. $\sum_{n=1}^{\infty} \left(\frac{1}{n} - \frac{1}{n^2}\right)^2$

23. $\sum_{n=1}^{\infty} \frac{1}{(\ln n)^2}$

24. $\sum_{n=1}^{\infty} \frac{n^2}{(\ln n)^n}$

In Exercises 25-34, determine the convergence of the given series. State the test used; more than one test may be appropriate.

25. $\sum_{n=1}^{\infty} \frac{n^2+4n-2}{n^3+4n^2-3n+7}$

26. $\sum_{n=1}^{\infty} \frac{n^4 4^n}{n!}$

27. $\sum_{n=1}^{\infty} \frac{n^2}{3^n+n}$

28. $\sum_{n=1}^{\infty} \frac{3^n}{n^n}$

29. $\sum_{n=1}^{\infty} \frac{n}{\sqrt{n^2+4n+1}}$

30. $\sum_{n=1}^{\infty} \frac{n!n!n!}{(3n)!}$

31. $\sum_{n=1}^{\infty} \frac{1}{\ln n}$

32. $\sum_{n=1}^{\infty} \left(\frac{n+2}{n+1}\right)$

33. $\sum\limits_{n=1}^{\infty} \frac{n^3}{(\ln n)^n}$

34. $\sum\limits_{n=1}^{\infty} \left(\frac{1}{n} - \frac{1}{n+2}\right)$

8.5: Alternating Series and Absolute Convergence

Terms and Concepts

1. Why is $\sum\limits_{n=1}^{\infty} \sin n$ not an alternating series?

2. A series $\sum\limits_{n=1}^{\infty} (-1)^n a_n$ converges when a_n is ______ ______ and $\lim\limits_{n\to\infty} a_n =$______.

3. Give an example of a series when $\sum\limits_{n=0}^{\infty} a_n$ converges but $\sum\limits_{n=0}^{\infty} |a_n|$ does not.

4. The sum of a ______ convergent series can be changed by rearranging the order of its terms.

Problems

In Exercises 5-20, an alternating series $\sum\limits_{n=i}^{\infty} a_n$ is given.

(a) Determine if the series converges or diverges.

(b) Determine if $\sum\limits_{n=0}^{\infty} |a_n|$ converges or diverges.

(c) If $\sum\limits_{n=0}^{\infty} a_n$ converges, determine if the convergence is conditional or absolute.

5. $\sum\limits_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^2}$

6. $\sum\limits_{n=1}^{\infty} \frac{(-1)^{n+1}}{\sqrt{n!}}$

7. $\sum\limits_{n=0}^{\infty} (-1)^n \frac{n+5}{3n-5}$

8. $\sum\limits_{n=1}^{\infty} (-1)^n \frac{2n}{n^2}$

9. $\sum\limits_{n=0}^{\infty} (-1)^{n+1} \frac{3n+5}{n^2-3n+1}$

10. $\sum\limits_{n=1}^{\infty} \frac{(-1)^n}{\ln n+1}$

11. $\sum\limits_{n=2}^{\infty} (-1)^n \frac{n}{\ln n}$

12. $\sum\limits_{n=1}^{\infty} \frac{(-1)^{n+1}}{1+3+5+\ldots+(2n-1)}$

13. $\sum\limits_{n=1}^{\infty} \cos(\pi n)$

14. $\sum\limits_{n=1}^{\infty} \frac{\sin((n+1/2)\pi)}{n \ln n}$

15. $\sum\limits_{n=0}^{\infty} \left(-\frac{2}{3}\right)^n$

16. $\sum\limits_{n=0}^{\infty} (-e)^{-n}$

17. $\sum\limits_{n=0}^{\infty} \frac{(-1)^n n^2}{n!}$

18. $\sum\limits_{n=0}^{\infty} (-1)^n 2^{-n^2}$

19. $\sum\limits_{n=0}^{\infty} \frac{(-1)^n}{\sqrt{n}}$

20. $\sum\limits_{n=1}^{\infty} \frac{(-1000)^2}{n!}$

Let S_n be the nth partial sum of a series. In Exercises 21-24, a convergent alternating series is given and a value of *n*. Compute S_n and S_{n+1} and use these values to find bounds on the sum of the series.

21. $\sum\limits_{n=1}^{\infty} \frac{(-1)^n}{\ln(n+1)}, \quad n=5$

22. $\sum\limits_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^4}, \quad n=4$

23. $\sum\limits_{n=0}^{\infty} \frac{(-1)^n}{n!}, \quad n=6$

24. $\sum\limits_{n=0}^{\infty} \left(-\frac{1}{2}\right)^n, \quad n=9$

In Exercises 25-28, a convergent alternating series is given along with its sum and a value of ϵ. Use Theorem 71 to find *n* such that the nth partial sum of the series is within ϵ of a sum of the series.

25. $\sum\limits_{n=0}^{\infty} \frac{(-1)^{n+1}}{n^4} = \frac{7\pi^4}{720}, \quad \epsilon = 0.001$

26. $\sum\limits_{n=0}^{\infty} \frac{(-1)^n}{n!} = \frac{1}{e}, \quad \epsilon = 0.0001$

27. $\sum\limits_{n=0}^{\infty} \frac{(-1)^n}{2n+1} = \frac{\pi}{4}, \quad \epsilon = 0.001$

28. $\sum\limits_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} = \cos 1, \quad \epsilon = 10^{-8}$

8.6: Power Series

Terms and Concepts

1. We adopt the convection that $x^0 =$______, regardless of the value of *x*.

2. What is the difference between the radius of convergence and the interval of convergence?

3. If the radius of convergence of $\sum\limits_{n=0}^{\infty} a_x x^n$ is 5, what is the radius of convergence of $\sum\limits_{n=1}^{\infty} n \cdot a_n x^{n-1}$?

4. If the radius of convergence of $\sum\limits_{n=0}^{\infty} a_n x^n$ is 5, what is the radius of convergence of $\sum\limits_{n=0}^{\infty} (=1)^n a_n x^n$?

Problems

In Exercises 5-8, write out the sum of the first 5 terms of the given power series.

5. $\sum\limits_{n=0}^{\infty} 2^n x^n$

6. $\sum\limits_{n=1}^{\infty} \frac{1}{n^2} x^n$

7. $\sum\limits_{n=0}^{\infty} \frac{1}{n!} x^n$

8. $\sum\limits_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n}$

In Exercises 9-24, a power series is given.
(a) Find the radius of convergence.
(b) Find the interval of convergence.

9. $\sum\limits_{n=0}^{\infty} \frac{(-1)^{n+1}}{n!} x^n$

10. $\sum\limits_{n=0}^{\infty} n x^n$

11. $\sum\limits_{n=1}^{\infty} \frac{(-1)^n (x-3)^n}{n}$

12. $\sum\limits_{n=0}^{\infty} \frac{(x+4)^n}{n!}$

13. $\sum_{n=0}^{\infty} \frac{x^n}{2^n}$

14. $\sum_{n=0}^{\infty} \frac{(-1)^n(x-5)^n}{10^n}$

15. $\sum_{n=0}^{\infty} 5^n(x-1)^n$

16. $\sum_{n=0}^{\infty} (-2)^n x^n$

17. $\sum_{n=0}^{\infty} \sqrt{n}x^n$

18. $\sum_{n=0}^{\infty} \frac{n}{3^n}x^n$

19. $\sum_{n=0}^{\infty} \frac{3^n}{n!}(x-5)^n$

20. $\sum_{n=0}^{\infty} (-1)^n n!(x-10)^n$

21. $\sum_{n=1}^{\infty} \frac{x^n}{n^2}$

22. $\sum_{n=1}^{\infty} \frac{(x+2)^n}{n^3}$

23. $\sum_{n=0}^{\infty} n!\left(\frac{x}{10}\right)^n$

24. $\sum_{n=0}^{\infty} n^2\left(\frac{x+4}{4}\right)^n$

In Exercises 25-30, a function $f(x) = \sum_{n=0}^{\infty} a_n x^n$ is given.
(a) Give a power series for $f'(x)$ and its interval of convergence.
(b) Give a power series for $f;(x)\,dx$ and its interval of convergence.

25. $\sum_{n=0}^{\infty} nx^n$

26. $\sum_{n=1}^{\infty} \frac{x^n}{n}$

27. $\sum_{n=0}^{\infty} \left(\frac{x}{2}\right)^n$

28. $\sum_{n=0}^{\infty} (-3x)^n$

29. $\sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}$

30. $\sum_{n=0}^{\infty} \frac{(-1)^n x^n}{n!}$

In Exercises 31-36, give the first 5 terms of the series that is a solution to the given differential equation.

31. $y' = 3y, \quad y(0) = 1$

32. $y' = 5y, \quad y(0) = 5$

33. $y' = y^2, \quad y(0) = 1$

34. $y' = y + 1, \quad y(0) = 1$

35. $y'' = -y, \quad y(0) = 0,\ y'(0) = 1$

36. $y'' = 2y, \quad y(0) = 1,\ y'(0) = 1$

8.7: Taylor Polynomials

Terms and Concepts

1. What is the difference between a Taylor polynomial and a Maclaurin polynomial?

2. T/F: In general, $p_n(x)$ approximates $f(x)$ better and better as *n* gets larger.

3. For some function $f(x)$, the Maclaurin polynomial of degree 3 is $p_4(x) = 6 + 3x - 4x^2 + 5x^3 - 7x^4$. What is $p_2(x)$?

4. For some function $f(x)$, the Maclaurin polynomial of degree 3 is $p_4(x) = 6 + 3x - 4x^2 + 5x^3 - 7x^4$. What is $f'''(0)$?

Problems

In Exercises 5-12, find the Maclaurin polynomial of degree *n* for the given function.

5. $f(x) = e^{-x}, \quad n = 3.$

6. $f(x) = \sin x, \quad n = 8.$

7. $f(x) = x \cdot e^x, \quad n = 5.$

8. $f(x) = \tan x, \quad n = 6.$

9. $f(x) = e^{2x}, \quad n = 4.$

10. $f(x) = \frac{1}{1-x}, \quad n = 4.$

11. $f(x) = \frac{1}{1+x}, \quad n = 4.$

12. $f(x) = \frac{1}{1-x}, \quad n = 7.$

In Exercises 13-20, find the Taylor polynomial of degree *n*, at $x = c$, for the given function.

13. $f(x) = \sqrt{x}, \quad n = 4, \quad c = 1$

14. $f(x) = \ln(x+1), \quad n = 4, \quad c = 1$

15. $f(x) = \cos x, \quad n = 6, \quad c = \pi/4$

16. $f(x) = \sin x, \quad n = 5, \quad c = \pi/6$

17. $f(x) = \frac{1}{x}, \quad n = 5, \quad c = 2$

18. $f(x) = \frac{1}{x^2}, \quad n = 8, \quad c = 1$

19. $f(x) = \frac{1}{x^2+1}, \quad n = 4, \quad c = -1$

20. $f(x) = x^2 \cos x, \quad n = 2, \quad c = -1$

In Exercises 21-24, approximate the function value with the indicated Taylor polynomial and give approximate bounds on the error.

21. Approximate$\sin 0.1$ with the Maclaurin polynomial of degree 3.

22. Approximate $\cos 1$ with the Maclaurin polynomial of degree 4.

23. Approximate $\sqrt{10}$ with the Taylor polynomial of degree 2 centered at $x = 9$.

24. Approximate $\ln 1.5$ with the Taylor polynomial of degree 3 centered at $x = 1$.

Exercises 25-28 ask for an *n* to be found such that $p_n(x)$ approximates $f(x)$ within a certain bound of accuracy.

25. Find *n* such that the Maclaurin polynomial of degree *n* of $f(x) = e^x$ approximates within 0.0001 of the actual value.

26. Find *n* such that the Taylor polynomial of degree *n* of $f(x) = \sqrt{x}$, centered at $x = 4$, approximates $\sqrt{3}$ within 0.0001 of the actual value.

27. Find *n* such that the Maclaurin polynomial of degree *n* of $f(x) = \cos x$ approximates $\cos \pi/3$ within 0.0001 of the actual value.

28. Find *n* such that the Maclaurin polynomial of degree *n* of $f(x) = \sin x$ approximates $\cos \pi$ within 0.0001 of the actual value.

In Exercises 29-33, find the n^{th} term of the indicated Taylor polynomial.

29. Find a formula for the n^{th} term of the Maclaurin polynomial for $f(x) = e^x$.

30. Find a formula for the n^{th} term of the Maclaurin polynomial for $f(x) = \cos x$.

31. Find a formula for the n^{th} term of the Maclaurin polynomial for $f(x) = \frac{1}{1-x}$.

32. Find a formula for the n^{th} term of the Maclaurin polynomial for $f(x) = \frac{1}{1+x}$.

33. Find a formula for the n^{th} term of the Maclaurin polynomial for $f(x) = \ln x$.

In Exercises 34-36, approximate the solution to the given differential equation with a degree 4 Maclaurin polynomial.

34. $y' = y, \quad y(0) = 1$

35. $y' = 5y, \quad y(0) = 3$

36. $y' = \frac{2}{y}, \quad y(0) = 1$

8.8: Taylor Series

Terms and Concepts

1. What is the difference between a Taylor polynomial and a Taylor series?

2 What theorem must we use to show that a function is equal to its Taylor series?

Problems

Key Idea 32 gives the n^{th} term of the Taylor series of common functions. In Exercises 3-6, verify the formula given in the Key Idea by finding the first few terms of the Taylor series of the given function and identifying a pattern.

3. $f(x) = e^x; \quad c = 0$

4. $f(x) = \sin x; \quad c = 0$

5. $f(x) = 1/(1-x); \quad c = 0$

6. $f(x) = \tan^{-1} x; \quad c = 0$

In Exercises 7-12, find a formula for the n^{th} term of the Taylor series of $f(x)$, centered at *c*, by finding the coefficients of the first few powers of *x* and looking for a pattern. (The formulas for several of these are found in Key Idea 32; show work verifying these formula.)

7. $f(x) = \cos x; \quad c = \pi/2$

8. $f(x) = 1/x; \quad c = 1$

9. $f(x) = e^{-x}; \quad c = \pi/2$

10. $f(x) = \ln(1+x); \quad c = 0$

11. $f(x) = x/(x+1); \quad c = 1$

12. $f(x) = \sin x; \quad c = \pi/4$

In Exercises 13-16, show that the Taylor series for $f(x)$, as given in Key Idea 32, is equal to $f(x)$ by applying Theorem 77; that is, show $\lim\limits_{n\to\infty} R_n(x) = 0$.

13. $f(x) = e^x$

14. $f(x) = \sin x$

15. $f(x) = \ln x$

16. $f(x) = 1/(1-x)$ (show equality only on (-1,0))

In Exercises 17-20, use the Taylor series given in Key Idea 32 to verify the given identity.

17. $\cos(-x) = \cos x$

18. $\sin(-x) = -\sin x$

19. $\frac{d}{dx}(\sin x) = \cos x$

20. $\frac{d}{dx}(\cos x) = -\sin x$

In Exercises 21-24, write out the first 5 terms of the Binomial series with the given k-value.

21. $k = 1/2$

22. $k = -1/2$

23. $k=1/3$

24. $k=4$

In Exercises 25-30, use the Taylor series given in Key Idea 32 to create the Taylor series of the given functions.

25. $f(x)=\cos\left(x^2\right)$

26. $f(x)=e^{-x}$

27. $f(x)=\sin(2x+3)$

28. $f(x)=\tan^{-1}(x/2)$

29. $f(x)=e^x\sin x$ (only find the first 4 terms).

30. $f(x)=(1+x)^{1/2}\cos x$ (only find the first 4 terms)

In Exercises 31-32, approximate the value of the given definite integral by using the first 4 nonzero terms of the integrand's Taylor series.

31. $\int_0^{\sqrt{\pi}}\sin\left(x^2\right)\,dx$

32.$\int_0^{\pi^2/4}\cos\left(\sqrt{x}\right)\,dx$

8.E: Applications of Sequences and Series (Exercises) is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

9: Curves in the Plane

We have explored functions of the form $y = f(x)$ closely throughout this text. We have explored their limits, their derivatives and their antiderivatives; we have learned to identify key features of their graphs, such as relative maxima and minima, inflection points and asymptotes; we have found equations of their tangent lines, the areas between portions of their graphs and the x-axis, and the volumes of solids generated by revolving portions of their graphs about a horizontal or vertical axis.

Despite all this, the graphs created by functions of the form $y = f(x)$ are limited. Since each x-value can correspond to only 1 y-value, common shapes like circles cannot be fully described by a function in this form. Fittingly, the “vertical line test” excludes vertical lines from being functions of x, even though these lines are important in mathematics.

In this chapter we’ll explore new ways of drawing curves in the plane. We’ll still work within the framework of functions, as an input will still only correspond to one output. However, our new techniques of drawing curves will render the vertical line test pointless, and allow us to create important – and beautiful – new curves. Once these curves are defined, we’ll apply the concepts of calculus to them, continuing to find equations of tangent lines and the areas of enclosed regions.

9.1: Conic Sections
9.2: Parametric Equations
9.3: Calculus and Parametric Equations
9.4: Introduction to Polar Coordinates
9.5: Calculus and Polar Functions
9.E: Applications of Curves in a Plane (Exercises)

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 9: Curves in the Plane is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

9.1: Conic Sections

The ancient Greeks recognized that interesting shapes can be formed by intersecting a plane with a *double napped* cone (i.e., two identical cones placed tip--to--tip as shown in the following figures). As these shapes are formed as sections of conics, they have earned the official name "conic sections."

The three "most interesting" conic sections are given in the top row of Figure $9.1.1$. They are the parabola, the ellipse (which includes circles) and the hyperbola. In each of these cases, the plane does not intersect the tips of the cones (usually taken to be the origin).

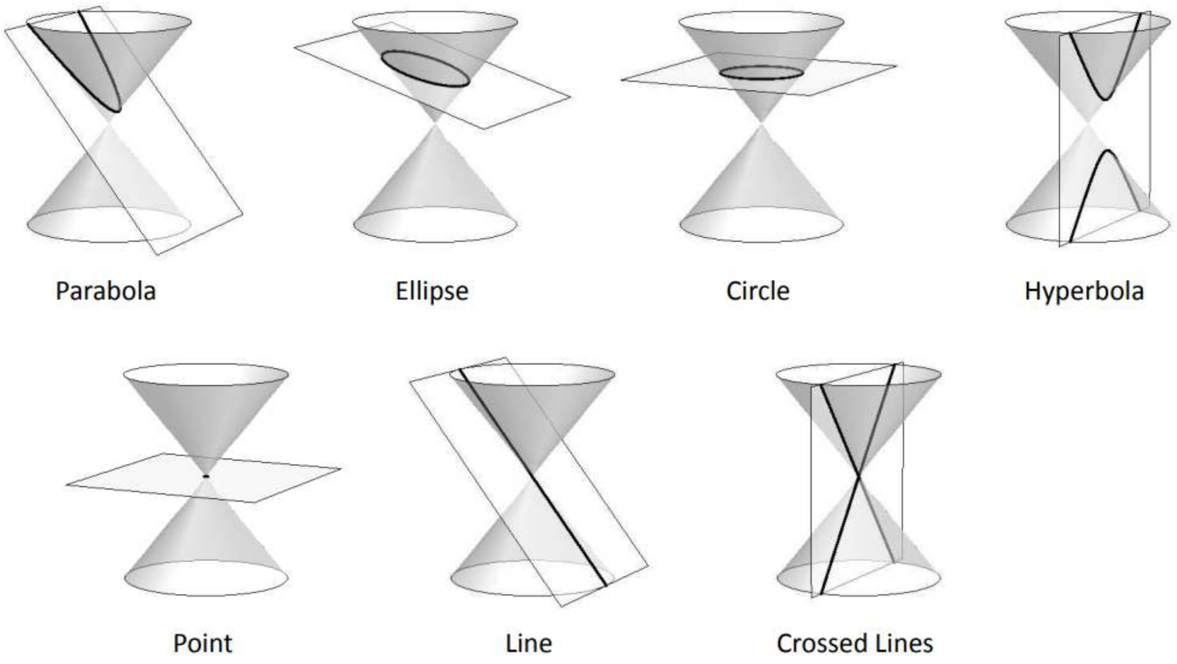


Figure $9.1.1$: Conic Sections

When the plane does contain the origin, three **degenerate** sections can be formed as shown the bottom row of Figure $9.1.1$: a point, a line, and crossed lines. We focus here on the nondegenerate cases.

While the above geometric constructs define the conics in an intuitive, visual way, these constructs are not very helpful when trying to analyze the shapes algebraically or consider them as the graph of a function. It can be shown that all conics can be defined by the general second--degree equation

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0. \tag{9.1.1}$$

While this algebraic definition has its uses, most find another geometric perspective of the conics more beneficial. Each nondegenerate conic can be defined as the **locus**, or set, of points that satisfy a certain distance property. These distance properties can be used to generate an algebraic formula, allowing us to study each conic as the graph of a function.

Large Parabolas

Definition 40: parabolas

A **parabola** is the locus of all points equidistant from a point (called a **focus**) and a line (called the **directrix**) that does not contain the focus.

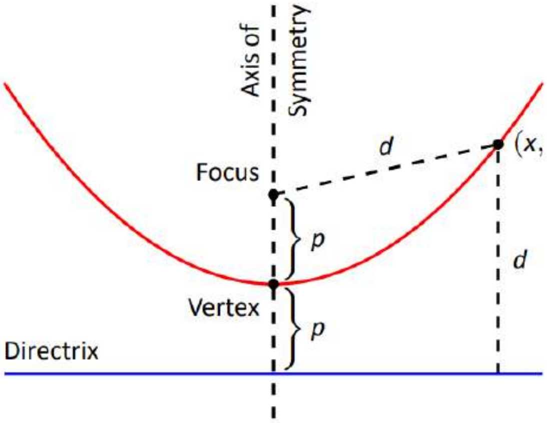


Figure 9.1.2: Illustrating the definition of the parabola and establishing an algebraic formula.

Figure 9.1.1 illustrates this definition. The point halfway between the focus and the directrix is the **vertex**. The line through the focus, perpendicular to the directrix, is the **axis of symmetry**, as the portion of the parabola on one side of this line is the mirror--image of the portion on the opposite side.

The definition leads us to an algebraic formula for the parabola. Let $P=(x,y)$ be a point on a parabola whose focus is at $F=(0,p)$ and whose directrix is at $y=-p$. (We'll assume for now that the focus lies on the y- axis; by placing the focus p units above the x-axis and the directrix p units below this axis, the vertex will be at $(0,0)$.)

We use the Distance Formula to find the distance d_1 between F and P:

$$d_1=\sqrt{(x-0)^2+(y-p)^2}. \tag{9.1.2}$$

The distance d_2 from P to the directrix is more straightforward:

$$d_2=y-(-p)=y+p. \tag{9.1.3}$$

These two distances are equal. Setting $d_1=d_2$, we can solve for y in terms of x:

$$\begin{aligned} d_1 &= d_2 \\ \sqrt{x^2+(y-p)^2} &= y+p \\ \text{Now square both sides.} & \\ x^2+(y-p)^2 &= (y+p)^2 \\ x^2+y^2-2yp+p^2 &= y^2+2yp+p^2 \\ x^2 &= 4yp \\ y &= \frac{1}{4p}x^2. \end{aligned}$$

The geometric definition of the parabola has led us to the familiar quadratic function whose graph is a parabola with vertex at the origin. When we allow the vertex to not be at $(0,0)$, we get the following standard form of the parabola.

key idea 33: general equation of a parabola

1. Vertical Axis of Symmetry: The equation of the parabola with vertex at (h,k) and directrix $y=k-p$ in standard form is

$$y=\frac{1}{4p}(x-h)^2+k. \tag{9.1.4}$$

   The focus is at $(h,k+p)$.
2. **Horizontal Axis of Symmetry**: The equation of the parabola with vertex at (h,k) and directrix $x=h-p$ in standard form is

$$x=\frac{1}{4p}(y-k)^2+h. \tag{9.1.5}$$

   The focus is at $(h+p,k)$.

Note: p is not necessarily a positive number.

Example 9.1.1: Finding the equation of a parabola

Give the equation of the parabola with focus at $(1,2)$ and directrix at $y=3$.

Solution

The vertex is located halfway between the focus and directrix, so $(h,k)=(1,2.5)$. This gives $p=-0.5$. Using Key Idea 33 we have the equation of the parabola as

$$y=\frac{1}{4(-0.5)}(x-1)^2+2.5=-\frac{1}{2}(x-1)^2+2.5. \tag{9.1.6}$$

The parabola is sketched in Figure 9.1.3.

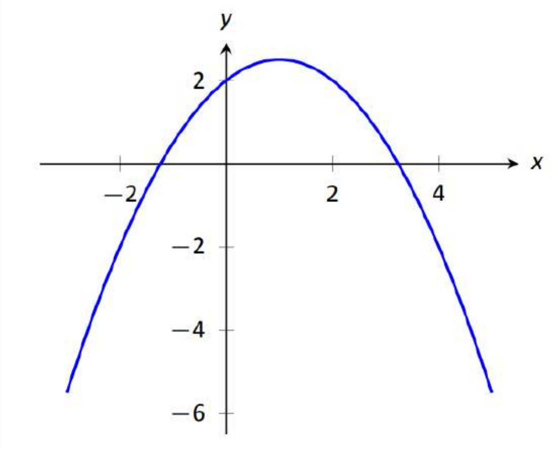


Figure 9.1.3: The parabola described in Example 9.1.1

Example 9.1.2: Finding the focus and directrix of a parabola

Find the focus and directrix of the parabola $x=\frac{1}{8}y^2-y+1$. The point $(7,12)$ lies on the graph of this parabola; verify that it is equidistant from the focus and directrix.

Solution

We need to put the equation of the parabola in its general form. This requires us to complete the square:

$$\begin{aligned}x&=\frac{1}{8}y^2-y+1\\&=\frac{1}{8}\left(y^2-8y+8\right)\\&=\frac{1}{8}\left(y^2-8y+16-16+8\right)\\&=\frac{1}{8}\left((y-4)^2-8\right)\\&=\frac{1}{8}(y-4)^2-1.\end{aligned}$$

Hence the vertex is located at $(-1,4)$. We have $\frac{1}{8}=\frac{1}{4p}$, so $p=2$. We conclude that the focus is located at $(1,4)$ and the directrix is $x=-3$. The parabola is graphed in Figure 9.1.4, along with its focus and directrix.

The point $(7,12)$ lies on the graph and is $7-(-3)=10$ units from the directrix. The distance from $(7,12)$ to the focus is:

$$\sqrt{(7-1)^2+(12-4)^2}=\sqrt{100}=10. \tag{9.1.7}$$

Indeed, the point on the parabola is equidistant from the focus and directrix.

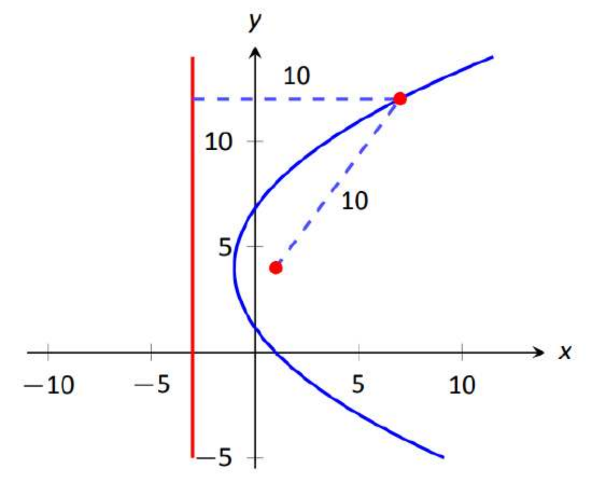


Figure $9.1.4$: The parabola described in Example $9.1.2$. The distances from a point on the parabola to the focus and directrix is given.

Reflective Property

One of the fascinating things about the nondegenerate conic sections is their reflective properties. Parabolas have the following reflective property:

Any ray emanating from the focus that intersects the parabola reflects off along a line perpendicular to the directrix.

This is illustrated in Figure $9.1.5$. The following theorem states this more rigorously.

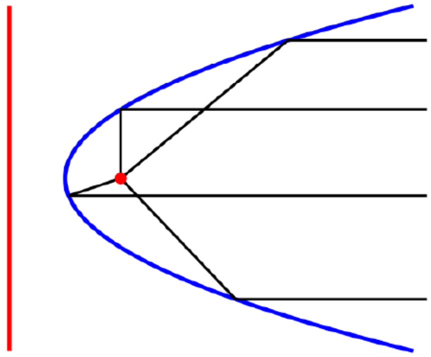


Figure $9.1.5$: Illustrating the parabola's reflective property.

THEOREM 79: REFLECTIVE PROPERTY OF THE PARABOLA

Let P be a point on a parabola. The tangent line to the parabola at P makes equal angles with the following two lines:

1. The line containing P and the focus F, and
2. The line perpendicular to the directrix through P.

Because of this reflective property, paraboloids (the 3D analogue of parabolas) make for useful flashlight reflectors as the light from the bulb, ideally located at the focus, is reflected along parallel rays. Satellite dishes also have paraboloid shapes. Signals coming from satellites effectively approach the dish along parallel rays. The dish then *focuses* these rays at the focus, where the sensor is located.

Ellipses

Definition 41: ellipse

An **ellipse** is the locus of all points whose sum of distances from two fixed points, each a **focus** of the ellipse, is constant.

An easy way to visualize this construction of an ellipse is to pin both ends of a string to a board. The pins become the foci. Holding a pencil tight against the string places the pencil on the ellipse; the sum of distances from the pencil to the pins is constant: the length of the string. See Figure $9.1.6$.

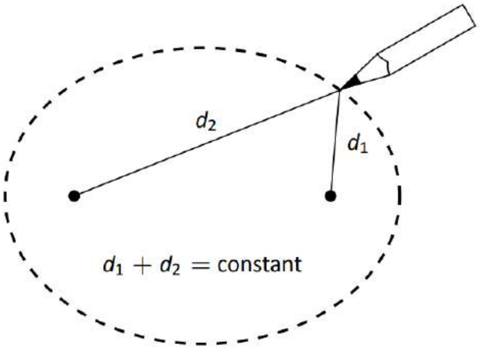


Figure $9.1.6$: Illustrating the construction of an ellipse with pins, pencil and string.

We can again find an algebraic equation for an ellipse using this geometric definition. Let the foci be located along the x- axis, c units from the origin. Let these foci be labeled as $F_1 = (-c, 0)$ and $F_2 = (c, 0)$. Let $P = (x, y)$ be a point on the ellipse. The sum of distances from F_1 to P (d_1) and from F_2 to P (d_2) is a constant d. That is, $d_1 + d_2 = d$. Using the Distance Formula, we have

$$\sqrt{(x+c)^2+y^2}+\sqrt{(x-c)^2+y^2}=d. \tag{9.1.8}$$

Using a fair amount of algebra can produce the following equation of an ellipse (note that the equation is an implicitly defined function; it has to be, as an ellipse fails the Vertical Line Test):

$$\frac{x^2}{\left(\frac{d}{2}\right)^2}+\frac{y^2}{\left(\frac{d}{2}\right)^2-c^2}=1. \tag{9.1.9}$$

This is not particularly illuminating, but by making the substitution $a = d/2$ and $b = \sqrt{a^2 - c^2}$, we can rewrite the above equation as

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1. \tag{9.1.10}$$

This choice of a and b is not without reason; as shown in Figure $9.1.7$, the values of a and b have geometric meaning in the graph of the ellipse.

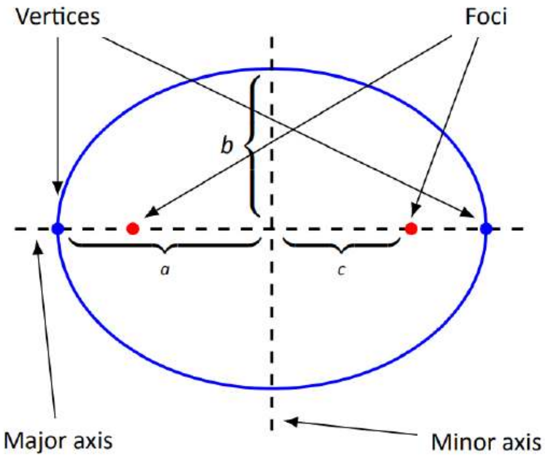


Figure $9.1.7$: Labeling the significant features of an ellipse.

In general, the two foci of an ellipse lie on the **major axis** of the ellipse, and the midpoint of the segment joining the two foci is the **center**. The major axis intersects the ellipse at two points, each of which is a **vertex**. The line segment through the center and perpendicular to the major axis is the **minor axis**. The "constant sum of distances" that defines the ellipse is the length of the major axis, i.e., $2a$.

Allowing for the shifting of the ellipse gives the following standard equations.

key idea 34: standard equation of the ellipse

The equation of an ellipse centered at (h,k) with major axis of length $2a$ and minor axis of length $2b$ in standard form is:

1. **Horizontal major axis**: $\frac{(x-h)^2}{a^2}+\frac{(y-k)^2}{b^2}=1$.
2. **Vertical major axis**: $\frac{(x-h)^2}{b^2}+\frac{(y-k)^2}{a^2}=1$.

The foci lie along the major axis, c units from the center, where $c^2=a^2-b^2$.

Example $9.1.3$: Finding the equation of an ellipse

Find the general equation of the ellipse graphed in Figure $9.1.8$.

Solution

The center is located at $(-3,1)$. The distance from the center to a vertex is 5 units, hence $a=5$. The minor axis seems to have length 4, so $b=2$. Thus the equation of the ellipse is

$$\frac{(x+3)^2}{4}+\frac{(y-1)^2}{25}=1. \tag{9.1.11}$$

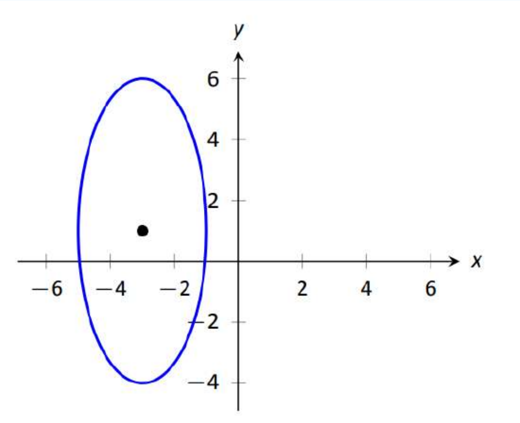


Figure $9.1.8$: The ellipse used in Example $9.1.3$.

Example $9.1.4$: Graphing an ellipse

Graph the ellipse defined by $4x^2+9y^2-8x-36y=-4$.

Solution

It is simple to graph an ellipse once it is in standard form. In order to put the given equation in standard form, we must complete the square with both the x and y terms. We first rewrite the equation by regrouping:

$$4x^2+9y^2-8x-36y=-4 \quad \Rightarrow \quad (4x^2-8x)+(9y^2-36y)=-4. \tag{9.1.12}$$

Now we complete the squares.

$$\begin{aligned}(4x^2-8x)+(9y^2-36y)&=-4\\4(x^2-2x)+9(y^2-4y)&=-4\\4(x^2-2x+1-1)+9(y^2-4y+4-4)&=-4\\4\big((x-1)^2-1\big)+9\big((y-2)^2-4\big)&=-4\\4(x-1)^2-4+9(y-2)^2-36&=-4\\4(x-1)^2+9(y-2)^2&=36\\\frac{(x-1)^2}{9}+\frac{(y-2)^2}{4}&=1.\end{aligned}$$

We see the center of the ellipse is at $(1,2)$. We have $a=3$ and $b=2$; the major axis is horizontal, so the vertices are located at $(-2,2)$ and $(4,2)$. We find $c=\sqrt{9-4}=\sqrt{5}\approx 2.24$. The foci are located along the major axis, approximately 2.24 units from the center, at $(1\pm 2.24,2)$. This is all graphed in Figure $9.1.9$.

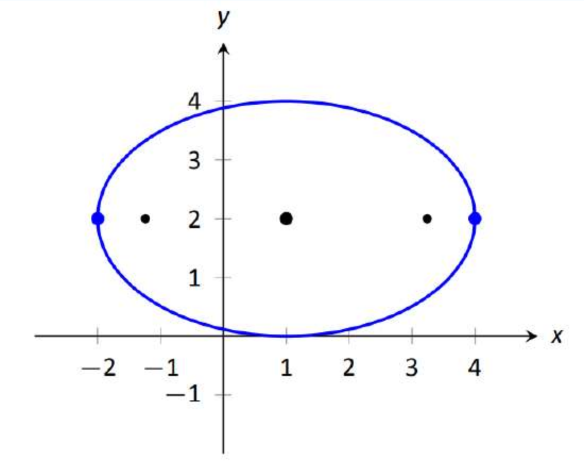


Figure $9.1.9$: Graphing the ellipse in Example $9.1.4$.

Eccentricity

When $a=b$, we have a circle. The general equation becomes

$$\frac{(x-h)^2}{a^2}+\frac{(y-k)^2}{a^2}=1 \quad\Rightarrow (x-h)^2+(y-k)^2=a^2, \tag{9.1.13}$$

the familiar equation of the circle centered at (h,k) with radius a. Since $a=b$, $c=\sqrt{a^2-b^2}=0$. The circle has "two" foci, but they lie on the same point, the center of the circle.

Consider Figure $9.1.1$, where several ellipses are graphed with $a=1$. In (a), we have $c=0$ and the ellipse is a circle. As c grows, the resulting ellipses look less and less circular. A measure of this "noncircularness" is *eccentricity*.

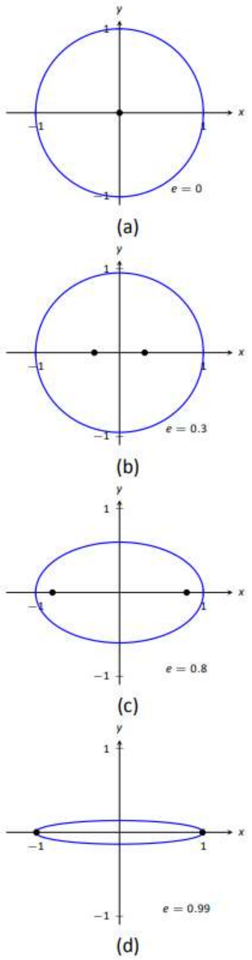


Figure $9.1.10$: Understanding the eccentricity of an ellipse.

Definition 42: eccentricity of an ellipse

The eccentricity e of an ellipse is $e=\frac{c}{a}$.

The eccentricity of a circle is 0; that is, a circle has no "noncircularness." As c approaches a, e approaches 1, giving rise to a very noncircular ellipse, as seen in Figure $9.1.10d$.

It was long assumed that planets had circular orbits. This is known to be incorrect; the orbits are elliptical. Earth has an eccentricity of 0.0167 -- it has a nearly circular orbit. Mercury's orbit is the most eccentric, with $e=0.2056$. (Pluto's eccentricity is greater, at $e=0.248$, the greatest of all the currently known dwarf planets.) The planet with the most circular orbit is Venus, with $e=0.0068$. The Earth's moon has an eccentricity of $e=0.0549$, also very circular.

Reflective Property

The ellipse also possesses an interesting reflective property. Any ray emanating from one focus of an ellipse reflects off the ellipse along a line through the other focus, as illustrated in Figure $9.1.11$. This property is given formally in the following theorem.

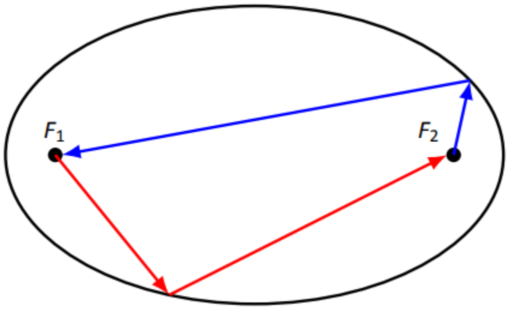


Figure 9.1.11: Illustrating the reflective property of an ellipse.

theorem 80: reflective property of an ellipse

Let P be a point on a ellipse with foci F_1 and F_2. The tangent line to the ellipse at P makes equal angles with the following two lines:

1. The line through F_1 and P, and
2. The line through F_2 and P.

This reflective property is useful in optics and is the basis of the phenomena experienced in whispering halls.

Hyperbolas

The definition of a hyperbola is very similar to the definition of an ellipse; we essentially just change the word "sum" to "difference."

Definition 43: hyperbola

A **hyperbola** is the locus of all points where the absolute value of difference of distances from two fixed points, each a focus of the hyperbola, is constant.

We do not have a convenient way of visualizing the construction of a hyperbola as we did for the ellipse. The geometric definition does allow us to find an algebraic expression that describes it. It will be useful to define some terms first.

The two foci lie on the **transverse axis** of the hyperbola; the midpoint of the line segment joining the foci is the **center** of the hyperbola. The transverse axis intersects the hyperbola at two points, each a vertex of the hyperbola. The line through the center and perpendicular to the transverse axis is the **conjugate axis**. This is illustrated in Figure 9.1.1. It is easy to show that the constant difference of distances used in the definition of the hyperbola is the distance between the vertices, i.e., $2a$.

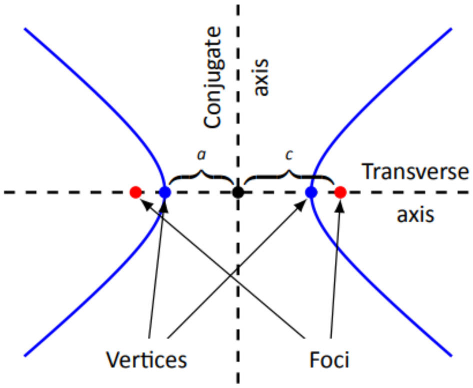


Figure 9.1.12: Labeling the significant features of a hyperbola.

key idea 35 standard equation of a hyperbola

The equation of a hyperbola centered at (h, k) in standard form is:

1. **Horizontal Transverse Axis:** $\frac{(x-h)^2}{a^2} - \frac{(y-k)^2}{b^2} = 1$.
2. **Vertical Transverse Axis:** $\frac{(y-k)^2}{a^2} - \frac{(x-h)^2}{b^2} = 1$.

The vertices are located a units from the center and the foci are located c units from the center, where $c^2 = a^2 + b^2$.

Graphing Hyperbolas

Consider the hyperbola $\frac{x^2}{9} - \frac{y^2}{1} = 1$. Solving for y, we find $y = \pm\sqrt{x^2/9 - 1}$. As x grows large, the "-1" part of the equation for y becomes less significant and $y \approx \pm\sqrt{x^2/9} = \pm x/3$. That is, as x gets large, the graph of the hyperbola looks very much like the lines $y = \pm x/3$. These lines are asymptotes of the hyperbola, as shown in Figure $9.1.13$.

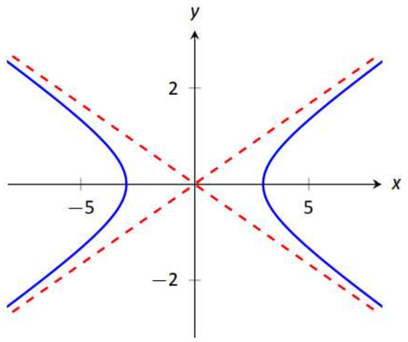


Figure $9.1.13$: Graphing the hyperbola $\frac{x^2}{9} - \frac{y^2}{1} = 1$ along with its asymptotes, $y = \pm x/3$.

This is a valuable tool in sketching. Given the equation of a hyperbola in general form, draw a rectangle centered at (h, k) with sides of length $2a$ parallel to the transverse axis and sides of length $2b$ parallel to the conjugate axis. (See Figure $9.1.14$ for an example with a horizontal transverse axis.) The diagonals of the rectangle lie on the asymptotes.

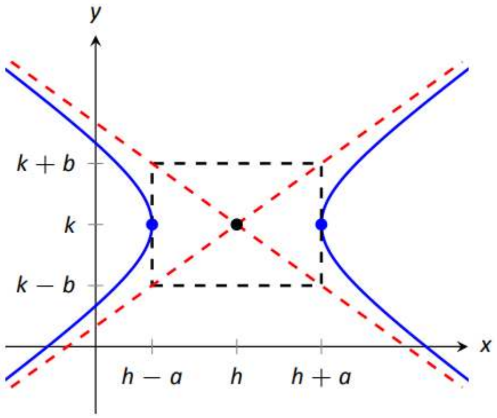


Figure $9.1.14$: Using the asymptotes of a hyperbola as a graphing aid.

These lines pass through (h, k). When the transverse axis is horizontal, the slopes are $\pm b/a$; when the transverse axis is vertical, their slopes are $\pm a/b$. This gives equations:

Horizontal Transverse Axis	Vertical Transverse Axis
$y = \pm\frac{b}{a}(x-h)+k$	$y = \pm\frac{a}{b}(x-h)+k.$

Example $9.1.5$: Graphing a hyperbola

Sketch the hyperbola given by $\frac{(y-2)^2}{25}-\frac{(x-1)^2}{4}=1$.

Solution

The hyperbola is centered at $(1,2)$; $a=5$ and $b=2$. In Figure $9.1.15$ we draw the prescribed rectangle centered at $(1,2)$ along with the asymptotes defined by its diagonals. The hyperbola has a vertical transverse axis, so the vertices are located at $(1,7)$ and $(1,-3)$. This is enough to make a good sketch.

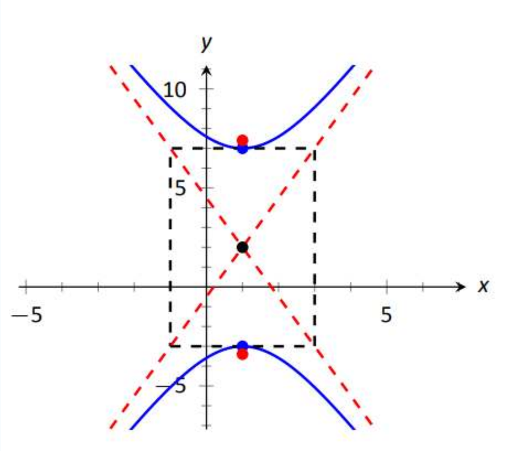


Figure $9.1.15$: Graphing the hyperbola in Example $9.1.5$.

We also find the location of the foci: as $c^2=a^2+b^2$, we have $c=\sqrt{29}\approx 5.4$. Thus the foci are located at $(1,2\pm 5.4)$ as shown in Figure $9.1.15$.

Example $9.1.6$: Graphing a hyperbola

Sketch the hyperbola given by $9x^2-y^2+2y=10$.

Solution

We must complete the square to put the equation in general form. (We recognize this as a hyperbola since it is a general quadratic equation and the x^2 and y^2 terms have opposite signs.)

$$\begin{aligned}9x^2-y^2+2y&=10\\9x^2-(y^2-2y)&=10\\9x^2-(y^2-2y+1-1)&=10\\9x^2-\big((y-1)^2-1\big)&=10\\9x^2-(y-1)^2&=9\\x^2-\frac{(y-1)^2}{9}&=1\end{aligned}$$

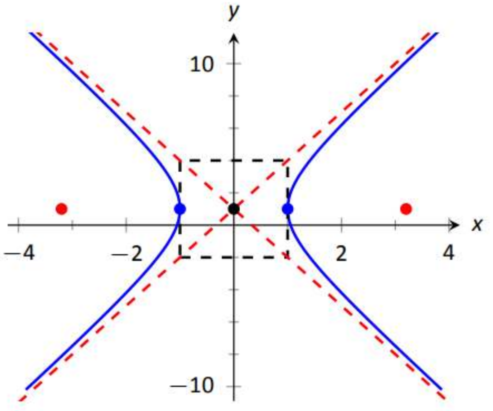


Figure $9.1.16$: Graphing the hyperbola in Example $9.1.6$.

We see the hyperbola is centered at $(0,1)$, with a horizontal transverse axis, where $a=1$ and $b=3$. The appropriate rectangle is sketched in Figure $9.1.16$ along with the asymptotes of the hyperbola. The vertices are located at $(\pm 1,1)$. We have $c=\sqrt{10}$ ≈ 3.2, so the foci are located at $(\pm 3.2,1)$ as shown in the figure.

Eccentricity

Definition 44: ECCENTRICITY OF A HYPERBOLA

The eccentricity of a hyperbola is $e=\frac{c}{a}$.

Note that this is the definition of eccentricity as used for the ellipse. When c is close in value to a (i.e., $e\approx 1$), the hyperbola is very narrow (looking almost like crossed lines). Figure $9.1.17$ shows hyperbolas centered at the origin with $a=1$. The graph in (a) has c $=1.05$, giving an eccentricity of $e=1.05$, which is close to 1. As c grows larger, the hyperbola widens and begins to look like parallel lines, as shown in Figure $9.1.17d$.

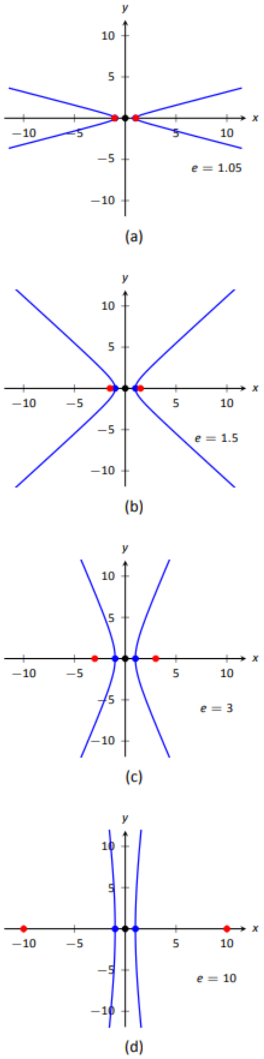


Figure 9.1.17: Understanding the eccentricity of a hyperbola.

Reflective Property

Hyperbolas share a similar reflective property with ellipses. However, in the case of a hyperbola, a ray emanating from a focus that intersects the hyperbola reflects along a line containing the other focus, but moving *away* from that focus. This is illustrated in Figure 9.1.19. Hyperbolic mirrors are commonly used in telescopes because of this reflective property. It is stated formally in the following theorem.

THEOREM 81: REFLECTIVE PROPERTY OF HYPERBOLAS

Let P be a point on a hyperbola with foci F_1 and F_2. The tangent line to the hyperbola at P makes equal angles with the following two lines:

1. The line through F_1 and P, and
2. The line through F_2 and P.

Location Determination

Determining the location of a known event has many practical uses (locating the epicenter of an earthquake, an airplane crash site, the position of the person speaking in a large room, etc.).

To determine the location of an earthquake's epicenter, seismologists use *trilateration* (not to be confused with *triangulation*). A seismograph allows one to determine how far away the epicenter was; using three separate readings, the location of the epicenter can be approximated.

A key to this method is knowing distances. What if this information is not available? Consider three microphones at positions A, B and C which all record a noise (a person's voice, an explosion, etc.) created at unknown location D. The microphone does not "know" when the sound was *created*, only when the sound was *detected*. How can the location be determined in such a situation?

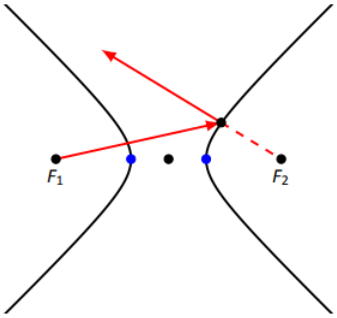


Figure $9.1.19$: Illustrating the reflective property of a hyperbola.

If each location has a clock set to the same time, hyperbolas can be used to determine the location. Suppose the microphone at position A records the sound at exactly 12:00, location B records the time exactly 1 second later, and location C records the noise exactly 2 seconds after that. We are interested in the *difference* of times. Since the speed of sound is approximately 340 m/s, we can conclude quickly that the sound was created 340 meters closer to position A than position B. If A and B are a known distance apart (as shown in Figure $9.1.1a$), then we can determine a hyperbola on which D must lie.

The "difference of distances" is 340; this is also the distance between vertices of the hyperbola. So we know $2a = 340$. Positions A and B lie on the foci, so $2c = 1000$. From this we can find $b \approx 470$ and can sketch the hyperbola, given in Figure $9.1.19b$. We only care about the side closest to A. (Why?)

We can also find the hyperbola defined by positions B and C. In this case, $2a = 680$ as the sound traveled an extra 2 seconds to get to C. We still have $2c = 1000$, centering this hyperbola at $(-500, 500)$. We find $b \approx 367$. This hyperbola is sketched in Figure $9.1.1c$. The intersection point of the two graphs is the location of the sound, at approximately $(188, -222.5)$.

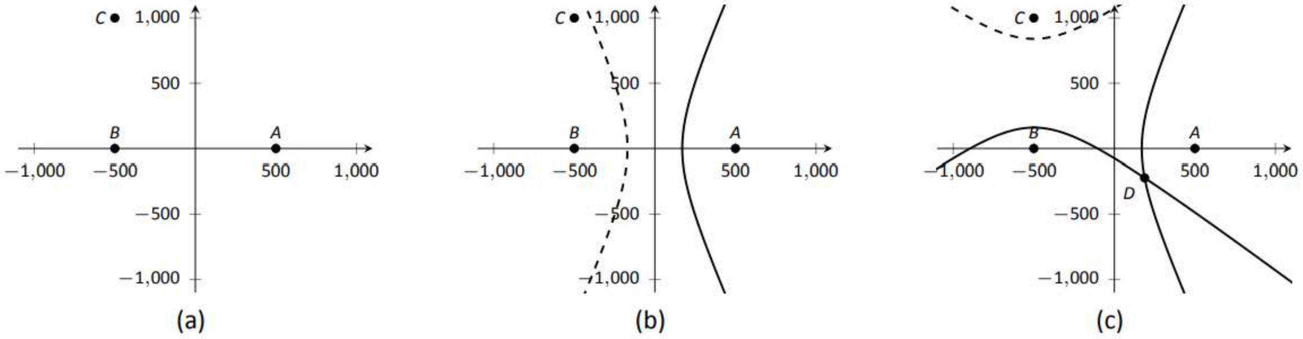


Figure $9.1.18$: Using hyperbolas in location detection.

This chapter explores curves in the plane, in particular curves that cannot be described by functions of the form $y = f(x)$. In this section, we learned of ellipses and hyperbolas that are defined implicitly, not explicitly. In the following sections, we will learn completely new ways of describing curves in the plane, using *parametric equations* and *polar coordinates*, then study these curves using calculus techniques.

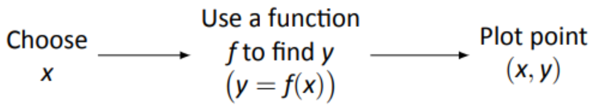
Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 9.1: Conic Sections is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

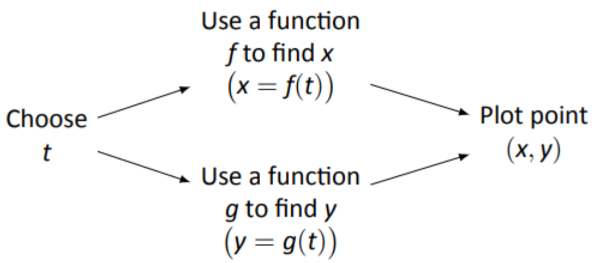
9.2: Parametric Equations

We are familiar with sketching shapes, such as parabolas, by following this basic procedure:



The **rectangular equation** $y = f(x)$ works well for some shapes like a parabola with a vertical axis of symmetry, but in the previous section we encountered several shapes that could not be sketched in this manner. (To plot an ellipse using the above procedure, we need to plot the "top" and "bottom" separately.)

In this section we introduce a new sketching procedure:



Here, x and y are found separately but then plotted together. This leads us to a definition.

Definition 45 Parametric Equations and Curves

Let f and g be continuous functions on an interval I. The set of all points $(x, y) = (f(t), g(t))$ in the Cartesian plane, as t varies over I, is the **graph** of the **parametric equations** $x = f(t)$ and $y = g(t)$, where t is the **parameter**. A **curve** is a graph along with the parametric equations that define it.

This is a formal definition of the word *curve*. When a curve lies in a plane (such as the Cartesian plane), it is often referred to as a **plane curve**. Examples will help us understand the concepts introduced in the definition.

Example $9.2.1$: Plotting parametric functions

Plot the graph of the parametric equations $x = t^2$, $y = t + 1$ for t in $[-2, 2]$.

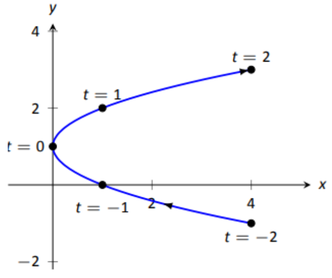
Solution

We plot the graphs of parametric equations in much the same manner as we plotted graphs of functions like $y = f(x)$: we make a table of values, plot points, then connect these points with a "reasonable" looking curve. Figure 9.20(a) shows such a table of values; note how we have 3 columns.

The points (x, y) from the table are plotted in Figure 9.20(b). The points have been connected with a smooth curve. Each point has been labeled with its corresponding t-value. These values, along with the two arrows along the curve, are used to indicate the **orientation** of the graph. This information helps us determine the direction in which the graph is "moving."

t	x	y
-2	4	-1
-1	1	0
0	0	1
1	1	2
2	4	3

(a)



(b)

Figure 9.20: A table of values of the parametric equations in Example 9.2.1 along with a sketch of their graph.

We often use the letter t as the parameter as we often regard t as representing *time*. Certainly there are many contexts in which the parameter is not time, but it can be helpful to think in terms of time as one makes sense of parametric plots and their orientation (for instance, "At time $t=0$ the position is $(1,2)$ and at time $t=3$ the position is $(5,1)$.").

Example $9.2.2$: Plotting parametric functions

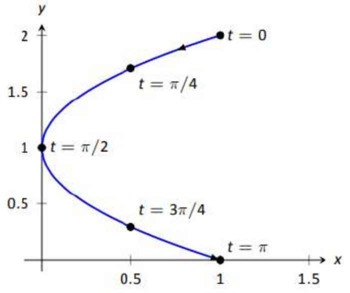
Sketch the graph of the parametric equations $x=\cos^2 t$, $y=\cos t+1$ for t in $[0,\pi]$.

Solution

We again start by making a table of values in Figure 9.21(a), then plot the points (x,y) on the Cartesian plane in Figure 9.21(b).

t	x	y
0	1	2
$\pi/4$	$1/2$	$1+\sqrt{2}/2$
$\pi/2$	0	1
$3\pi/4$	$1/2$	$1-\sqrt{2}/2$
π	1	0

(a)



(b)

Figure 9.21: A table of values of the parametric equations in Example 9.2.2 along with a sketch of their graph.

It is not difficult to show that the curves in Examples $9.2.1$ and $9.2.2$ are portions of the same parabola. While the *parabola* is the same, the *curves* are different. In Example $9.2.1$, if we let t vary over all real numbers, we'd obtain the entire parabola. In this example, letting t vary over all real numbers would still produce the same graph; this portion of the parabola would be traced, and re--traced, infinitely. The orientation shown in Figure 9.21 shows the orientation on $[0,\pi]$, but this orientation is reversed on $[\pi,2\pi]$.

These examples begin to illustrate the powerful nature of parametric equations. Their graphs are far more diverse than the graphs of functions produced by "$y=f(x)$" functions.

Technology Note: Most graphing utilities can graph functions given in parametric form. Often the word "parametric" is abbreviated as "PAR" or "PARAM" in the options. The user usually needs to determine the graphing window (i.e, the minimum and maximum x- and y-values), along with the values of t that are to be plotted. The user is often prompted to give a t minimum, a t maximum, and a "t-step" or "Δt." Graphing utilities effectively plot parametric functions just as we've shown here: they plots lots of points. A smaller t-step plots more points, making for a smoother graph (but may take longer). In Figure 9.20, the t-step is 1; in Figure 9.21, the t-step is $\pi/4$.

One nice feature of parametric equations is that their graphs are easy to shift. While this is not too difficult in the "$y = f(x)$" context, the resulting function can look rather messy. (Plus, to shift to the right by two, we replace x with $x-2$, which is counter--intuitive.) The following example demonstrates this.

Example 9.2.3: Shifting the graph of parametric functions

Sketch the graph of the parametric equations $x = t^2 + t$, $y = t^2 - t$. Find new parametric equations that shift this graph to the right 3 places and down 2.

Solution
The graph of the parametric equations is given in Figure 9.22 (a). It is a parabola with a axis of symmetry along the line $y = x$; the vertex is at $(0,0)$.

In order to shift the graph to the right 3 units, we need to increase the x-value by 3 for every point. The straightforward way to accomplish this is simply to add 3 to the function defining x: $x = t^2 + t + 3$. To shift the graph down by 2 units, we wish to decrease each y-value by 2, so we subtract 2 from the function defining y: $y = t^2 - t - 2$. Thus our parametric equations for the shifted graph are $x = t^2 + t + 3$, $y = t^2 - t - 2$. This is graphed in Figure 9.22 (b). Notice how the vertex is now at $(3,-2)$.

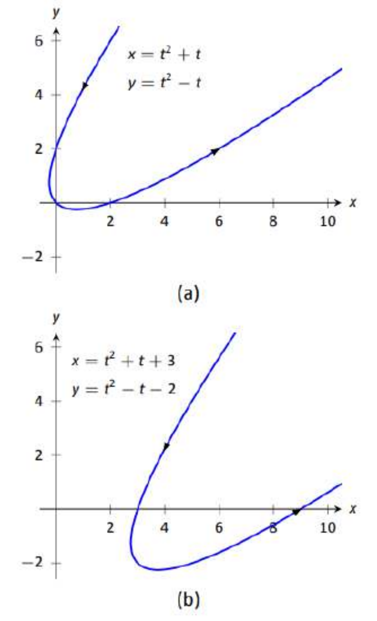


Figure 9.22: Illustrating how to shift graphs in 9.2.3.

Because the x- and y-values of a graph are determined independently, the graphs of parametric functions often possess features not seen on "$y = f(x)$" type graphs. The next example demonstrates how such graphs can arrive at the same point more than once.

Example 9.2.4: Graphs that cross themselves

Plot the parametric functions $x = t^3 - 5t^2 + 3t + 11$ and $y = t^2 - 2t + 3$ and determine the t-values where the graph crosses itself.

Solution
Using the methods developed in this section, we again plot points and graph the parametric equations as shown in Figure 9.23. It appears that the graph crosses itself at the point $(2,6)$, but we'll need to analytically determine this.

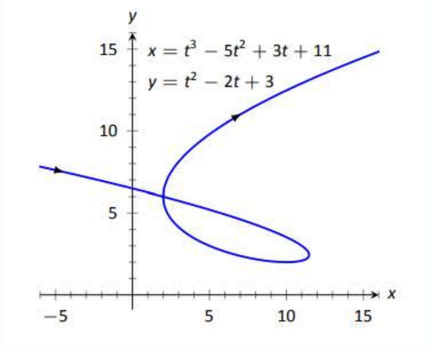


Figure 9.23: A graph of the parametric equations from Example 9.2.4.

We are looking for two different values, say, s and t, where $x(s) = x(t)$ and $y(s) = y(t)$. That is, the x-values are the same precisely when the y-values are the same. This gives us a system of 2 equations with 2 unknowns:

$$\begin{gathered} s^3 - 5s^2 + 3s + 11 = t^3 - 5t^2 + 3t + 11 \\ s^2 - 2s + 3 = t^2 - 2t + 3 \end{gathered} \tag{9.2.1}$$

Solving this system is not trivial but involves only algebra. Using the quadratic formula, one can solve for t in the second equation and find that $t = 1 \pm \sqrt{s^2 - 2s + 1}$. This can be substituted into the first equation, revealing that the graph crosses itself at $t = -1$ and $t = 3$. We confirm our result by computing $x(-1) = x(3) = 2$ and $y(-1) = y(3) = 6$.

Converting between Rectangular and Parametric Equations

It is sometimes useful to rewrite equations in rectangular form (i.e., $y = f(x)$) into parametric form, and vice--versa. Converting from rectangular to parametric can be very simple: given $y = f(x)$, the parametric equations $x = t$, $y = f(t)$ produce the same graph. As an example, given $y = x^2$, the parametric equations $x = t$, $y = t^2$ produce the familiar parabola. However, other parametrizations can be used. The following example demonstrates one possible alternative.

Example $9.2.5$: Converting from rectangular to parametric

Consider $y = x^2$. Find parametric equations $x = f(t)$, $y = g(t)$ for the parabola where $t = \frac{dy}{dx}$. That is, $t = a$ corresponds to the point on the graph whose tangent line has slope a.

Solution

We start by computing $\frac{dy}{dx}$: $y' = 2x$. Thus we set $t = 2x$. We can solve for x and find $x = t/2$. Knowing that $y = x^2$, we have $y = t^2/4$. Thus parametric equations for the parabola $y = x^2$ are

$$x = t/2 \quad y = t^2/4. \tag{9.2.2}$$

To find the point where the tangent line has a slope of -2, we set $t = -2$. This gives the point $(-1, 1)$. We can verify that the slope of the line tangent to the curve at this point indeed has a slope of -2.

We sometimes chose the parameter to accurately model physical behavior.

Example $9.2.6$: Converting from rectangular to parametric

An object is fired from a height of 0ft and lands 6 seconds later, 192ft away. Assuming ideal projectile motion, the height, in feet, of the object can be described by $h(x) = -x^2/64 + 3x$, where x is the distance in feet from the initial location. (Thus $h(0) = h(192) = 0\text{ft}$.) Find parametric equations $x = f(t)$, $y = g(t)$ for the path of the projectile where x is the horizontal distance the object has traveled at time t (in seconds) and y is the height at time t.

Solution

Physics tells us that the horizontal motion of the projectile is linear; that is, the horizontal speed of the projectile is constant. Since the object travels 192ft in 6s, we deduce that the object is moving horizontally at a rate of 32ft/s, giving the equation $x = 32t$. As $y = -x^2/64 + 3x$, we find $y = -16t^2 + 96t$. We can quickly verify that $y''' = -32\text{ft/s}^2$, the acceleration due to gravity, and that the projectile reaches its maximum at $t = 3$, halfway along its path.

These parametric equations make certain determinations about the object's location easy: 2 seconds into the flight the object is at the point $\big(x(2), y(2)\big) = \big(64, 128\big)$. That is, it has traveled horizontally 64ft and is at a height of 128ft, as shown in Figure 9.24.

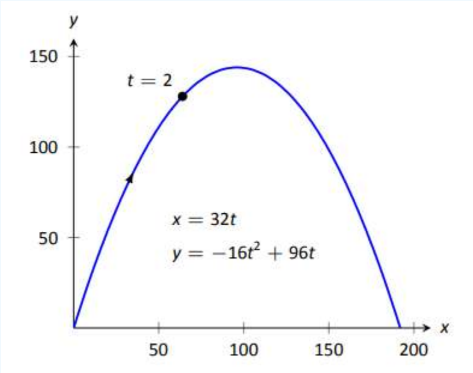


Figure 9.24: Graphing projectile motion in Example 9.2.6

It is sometimes necessary to convert given parametric equations into rectangular form. This can be decidedly more difficult, as some "simple" looking parametric equations can have very "complicated" rectangular equations. This conversion is often referred to as "eliminating the parameter," as we are looking for a relationship between x and y that does not involve the parameter t.

Example 9.2.7: Eliminating the parameter

Find a rectangular equation for the curve described by

$$x = \frac{1}{t^2+1} \quad \text{and} \quad y = \frac{t^2}{t^2+1}. \tag{9.2.3}$$

Solution

There is not a set way to eliminate a parameter. One method is to solve for t in one equation and then substitute that value in the second. We use that technique here, then show a second, simpler method.

Starting with $x = 1/(t^2+1)$, solve for t: $t = \pm\sqrt{1/x - 1}$. Substitute this value for t in the equation for y:

$$\begin{aligned} y &= \frac{t^2}{t^2+1} \\ &= \frac{1/x - 1}{1/x - 1 + 1} \\ &= \frac{1/x - 1}{1/x} \\ &= \left(\frac{1}{x} - 1\right) \cdot x \\ &= 1 - x. \end{aligned}$$

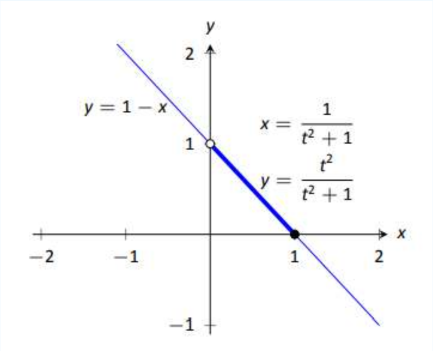


Figure 9.25: Graphing parametric and rectangular equations for a graph in Example 9.2.7

Thus $y = 1 - x$. One may have recognized this earlier by manipulating the equation for y:

$$y = \frac{\quad}{t^2+1} = 1 - \frac{\quad}{t^2+1} = 1 - x. \tag{9.2.4}$$

This is a shortcut that is very specific to this problem; sometimes shortcuts exist and are worth looking for.

We should be careful to limit the domain of the function $y = 1 - x$. The parametric equations limit x to values in $(0, 1]$, thus to produce the same graph we should limit the domain of $y = 1 - x$ to the same.

The graphs of these functions is given in Figure 9.25. The portion of the graph defined by the parametric equations is given in a thick line; the graph defined by $y = 1 - x$ with unrestricted domain is given in a thin line.

Example 9.2.8: Eliminating the parameter

Eliminate the parameter in $x = 4\cos t + 3$, $y = 2\sin t + 1$

Solution

We should not try to solve for t in this situation as the resulting algebra/trig would be messy. Rather, we solve for $\cos t$ and $\sin t$ in each equation, respectively. This gives

$$\cos t = \frac{x-3}{4} \quad \text{and} \quad \sin t = \frac{y-1}{2}. \tag{9.2.5}$$

The Pythagorean Theorem gives $\cos^2 t + \sin^2 t = 1$, so:

$$\begin{aligned} \cos^2 t + \sin^2 t &= 1 \\ \left(\frac{x-3}{4}\right)^2 + \left(\frac{y-1}{2}\right)^2 &= 1 \\ \frac{(x-3)^2}{16} + \frac{(y-1)^2}{4} &= 1 \end{aligned}$$

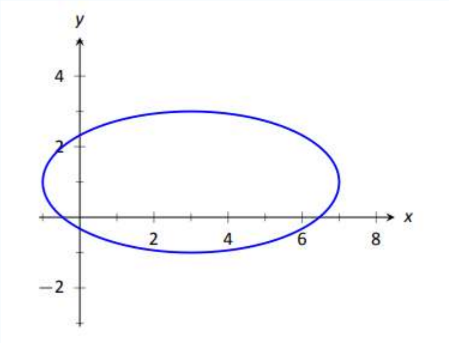


Figure 9.26: Graphing the parametric equations $x = 4\cos t + 3$, $y = 2\sin t + 1$ in Example 9.2.8

This final equation should look familiar -- it is the equation of an ellipse! Figure 9.26 plots the parametric equations, demonstrating that the graph is indeed of an ellipse with a horizontal major axis and center at $(3, 1)$.

The Pythagorean Theorem can also be used to identify parametric equations for hyperbolas. We give the parametric equations for ellipses and hyperbolas in the following Key Idea.

KEY IDEA 36 PARAMETRIC EQUATIONS OF ELLIPSES AND HYPERBOLAS

- The parametric equations

  $$x = a\cos t + h, \quad y = b\sin t + k \tag{9.2.6}$$

  define an ellipse with horizontal axis of length $2a$ and vertical axis of length $2b$, centered at (h, k).
- The parametric equations

  $$x = a\tan t + h, \quad y = \pm b\sec t + k \tag{9.2.7}$$

define a hyperbola with vertical transverse axis centered at (h, k), and

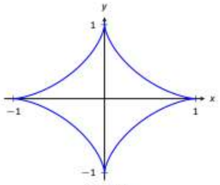
$$x = \pm a\sec t + h, \quad y = b\tan t + k \tag{9.2.8}$$

defines a hyperbola with horizontal transverse axis. Each has asymptotes at $y = \pm b/a(x-h)+k$.

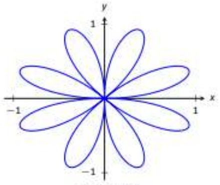
Special Curves

Figure 9.27 gives a small gallery of "interesting" and "famous" curves along with parametric equations that produce them. Interested readers can begin learning more about these curves through internet searches.

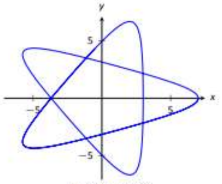
One might note a feature shared by two of these graphs: "sharp corners," or **cusps**. We have seen graphs with cusps before and determined that such functions are not differentiable at these points. This leads us to a definition.



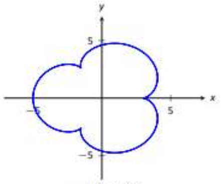
Astroid
$x = \cos^3 t$
$y = \sin^3 t$



Rose Curve
$x = \cos(t)\sin(4t)$
$y = \sin(t)\sin(4t)$



Hypotrochoid
$x = 2\cos(t) + 5\cos(2t/3)$
$y = 2\sin(t) - 5\sin(2t/3)$



Epicycloid
$x = 4\cos(t) - \cos(4t)$
$y = 4\sin(t) - \sin(4t)$

Figure 9.27: A gallery of interesting planar curves.

Definition 46 SMOOTH

A curve C defined by $x = f(t)$, $y = g(t)$ is **smooth** on an interval I if f' and g' are continuous on I and not simultaneously 0 (except possibly at the endpoints of I). A curve is **piecewise smooth** on I if I can be partitioned into subintervals where C is smooth on each subinterval.

Consider the astroid, given by $x = \cos^3 t$, $y = \sin^3 t$. Taking derivatives, we have:

$$x' = -3\cos^2 t\sin t \quad \text{and} \quad y' = 3\sin^2 t\cos t. \tag{9.2.9}$$

It is clear that each is 0 when $t = 0, \ \pi/2, \ \pi, \ldots$. Thus the astroid is not smooth at these points, corresponding to the cusps seen in the figure. We demonstrate this once more.

Example $9.2.9$: Determine where a curve is not smooth

Let a curve C be defined by the parametric equations $x = t^3 - 12t + 17$ and $y = t^2 - 4t + 8$. Determine the points, if any, where it is not smooth.

Solution

We begin by taking derivatives.

$$x' = 3t^2 - 12, \quad y' = 2t - 4. \tag{9.2.10}$$

We set each equal to 0:

$$\begin{aligned} x' = 0 \Rightarrow 3t^2 - 12 = 0 \Rightarrow t = \pm 2 \\ y' = 0 \Rightarrow 2t - 4 = 0 \Rightarrow t = 2 \end{aligned} \tag{9.2.11}$$

We see at $t = 2$ both x' and y' are 0; thus C is not smooth at $t = 2$, corresponding to the point $(1, 4)$. The curve is graphed in Figure 9.28, illustrating the cusp at $(1, 4)$.

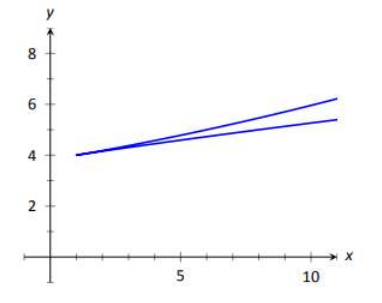


Figure 9.28: Graphing the curve in Example 9.2.9; note it is not smooth at $(1, 4)$.

If a curve is not smooth at $t = t_0$, it means that $x'(t_0) = y'(t_0) = 0$ as defined. This, in turn, means that rate of change of x (and y) is 0; that is, at that instant, neither x nor y is changing. If the parametric equations describe the path of some object, this means the object is at rest at t_0. An object at rest can make a "sharp" change in direction, whereas moving objects tend to change direction in a "smooth" fashion.

One should be careful to note that a "sharp corner" does not have to occur when a curve is not smooth. For instance, one can verify that $x = t^3$, $y = t^6$ produce the familiar $y = x^2$ parabola. However, in this parametrization, the curve is not smooth. A particle traveling along the parabola according to the given parametric equations comes to rest at $t = 0$, though no sharp point is created.\

Our previous experience with cusps taught us that a function was not differentiable at a cusp. This can lead us to wonder about derivatives in the context of parametric equations and the application of other calculus concepts. Given a curve defined parametrically, how do we find the slopes of tangent lines? Can we determine concavity? We explore these concepts and more in the next section.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 9.2: Parametric Equations is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

9.3: Calculus and Parametric Equations

The previous section defined curves based on parametric equations. In this section we'll employ the techniques of calculus to study these curves. We are still interested in lines tangent to points on a curve. They describe how the y-values are changing with respect to the x-values, they are useful in making approximations, and they indicate instantaneous direction of travel.

The slope of the tangent line is still $\frac{dy}{dx}$, and the Chain Rule allows us to calculate this in the context of parametric equations. If $x = f(t)$ and $y = g(t)$, the Chain Rule states that

$$\frac{dy}{dt} = \frac{dy}{dx} \cdot \frac{dx}{dt}. \tag{9.3.1}$$

Solving for $\frac{dy}{dx}$, we get

$$\frac{dy}{dx} = \frac{dy}{dt} \bigg/ \frac{dx}{dt} = \frac{g'(t)}{f'(t)}, \tag{9.3.2}$$

provided that $f'(t) \neq 0$. This is important so we label it a Key Idea.

key idea 37 Finding $\frac{dy}{dx}$ with Parametric Equations.

Let $x = f(t)$ and $y = g(t)$, where f and g are differentiable on some open interval I and $f'(t) \neq 0$ on I. Then

$$\frac{dy}{dx} = \frac{g'(t)}{f'(t)}. \tag{9.3.3}$$

We use this to define the tangent line.

Definition 47 Tangent and Normal Lines

Let a curve C be parametrized by $x = f(t)$ and $y = g(t)$, where f and g are differentiable functions on some interval I containing $t = t_0$. The **tangent line** to C at $t = t_0$ is the line through $\big(f(t_0), g(t_0)\big)$ with slope $m = g'(t_0)/f'(t_0)$, provided $f'(t_0) \neq 0$.

The **normal line** to C at $t = t_0$ is the line through $\big(f(t_0), g(t_0)\big)$ with slope $m = -f'(t_0)/g'(t_0)$, provided $g'(t_0) \neq 0$.

The definition leaves two special cases to consider. When the tangent line is horizontal, the normal line is undefined by the above definition as $g'(t_0) = 0$. Likewise, when the normal line is horizontal, the tangent line is undefined. It seems reasonable that these lines be defined (one can draw a line tangent to the "right side" of a circle, for instance), so we add the following to the above definition.

1. If the tangent line at $t = t_0$ has a slope of 0, the normal line to C at $t = t_0$ is the line $x = f(t_0)$.
2. If the normal line at $t = t_0$ has a slope of 0, the tangent line to C at $t = t_0$ is the line $x = f(t_0)$.

Example $9.3.1$: Tangent and Normal Lines to Curves

Let $x = 5t^2 - 6t + 4$ and $y = t^2 + 6t - 1$, and let C be the curve defined by these equations.

1. Find the equations of the tangent and normal lines to C at $t = 3$.
2. Find where C has vertical and horizontal tangent lines.

Solution

1. We start by computing $f'(t) = 10t - 6$ and $g'(t) = 2t + 6$. Thus

$$\frac{dy}{dx} = \frac{2t + 6}{10t - 6}. \tag{9.3.4}$$

Make note of something that might seem unusual: $\frac{dy}{dx}$ is a function of t, not x. Just as points on the curve are found in terms of t, so are the slopes of the tangent lines.

The point on C at $t = 3$ is $(31, 26)$. The slope of the tangent line is $m = 1/2$ and the slope of the normal line is $m = -2$. Thus,

- the equation of the tangent line is $y = \frac{1}{2}(x-31)+26$, and
- the equation of the normal line is $y = -2(x-31)+26$.

This is illustrated in Figure 9.29.

2. To find where C has a horizontal tangent line, we set $\frac{dy}{dx} = 0$ and solve for t. In this case, this amounts to setting $g'(t) = 0$ and solving for t (and making sure that $f'(t) \neq 0$).

$$g'(t) = 0 \quad \Rightarrow \quad 2t+6=0 \quad \Rightarrow \quad t=-3. \tag{9.3.5}$$

The point on C corresponding to $t=-3$ is $(67,-10)$; the tangent line at that point is horizontal (hence with equation $y = -10$).

To find where C has a vertical tangent line, we find where it has a horizontal normal line, and set $-\frac{f'(t)}{g'(t)} = 0$. This amounts to setting $f'(t)=0$ and solving for t (and making sure that $g'(t) \neq 0$).

$$f'(t) = 0 \quad \Rightarrow \quad 10t-6=0 \quad \Rightarrow \quad t=0.6. \tag{9.3.6}$$

The point on C corresponding to $t=0.6$ is $(2.2, 2.96)$. The tangent line at that point is $x = 2.2$.

The points where the tangent lines are vertical and horizontal are indicated on the graph in Figure 9.29.

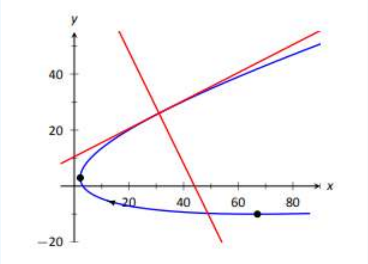


Figure 9.29: Graphing tangent and normal lines in Example 9.3.1

Example $9.3.2$: Tangent and Normal Lines to a Circle

1. Find where the unit circle, defined by $x = \cos t$ and $y = \sin t$ on $[0, 2\pi]$, has vertical and horizontal tangent lines.
2. Find the equation of the normal line at $t = t_0$.

Solution

1. We compute the derivative following Key Idea 37:

$$\frac{dy}{dx} = \frac{g'(t)}{f'(t)} = -\frac{\cos t}{\sin t}. \tag{9.3.7}$$

The derivative is 0 when $\cos t = 0$; that is, when $t = \pi/2,\ 3\pi/2$. These are the points $(0,1)$ and $(0,-1)$ on the circle.

The normal line is horizontal (and hence, the tangent line is vertical) when $\sin t = 0$; that is, when $t = 0,\ \pi,\ 2\pi$, corresponding to the points $(-1,0)$ and $(0,1)$ on the circle. These results should make intuitive sense.

2. The slope of the normal line at $t = t_0$ is $m = \frac{\sin t_0}{\cos t_0} = \tan t_0$. This normal line goes through the point $(\cos t_0, \sin t_0)$, giving the line

$$\begin{aligned} y &= \frac{\sin t_0}{\cos t_0}(x - \cos t_0) + \sin t_0 \\ &= (\tan t_0)x, \end{aligned}$$

as long as $\cos t_0 \neq 0$. It is an important fact to recognize that the normal lines to a circle pass through its center, as illustrated in Figure 9.30. Stated in another way, any line that passes through the center of a circle intersects the circle at right angles.

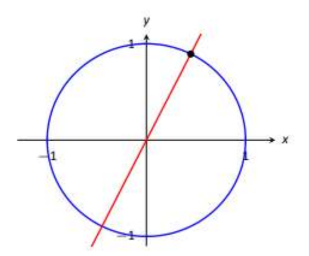


Figure 9.30: Illustrating how a circle's normal lines pass through its center.

Example 9.3.3: Tangent lines when $\frac{dy}{dx}$ is not defined

Find the equation of the tangent line to the astroid $x = \cos^3 t$, $y = \sin^3 t$ at $t = 0$, shown in Figure 9.31.

Solution
We start by finding $x'(t)$ and $y'(t)$:

$$x'(t) = -3\sin t\cos^2 t, \qquad y'(t) = 3\cos t\sin^2 t. \tag{9.3.8}$$

Note that both of these are 0 at $t = 0$; the curve is not smooth at $t = 0$ forming a cusp on the graph. Evaluating $\frac{dy}{dx}$ at this point returns the indeterminate form of "0/0".

We can, however, examine the slopes of tangent lines near $t = 0$, and take the limit as $t \to 0$.

$$\begin{aligned}\lim_{t\to 0}\frac{y'(t)}{x'(t)} &= \lim_{t\to 0}\frac{3\cos t\sin^2 t}{-3\sin t\cos^2 t} \qquad \text{(We can cancel as } t \neq 0.\text{)}\\ &= \lim_{t\to 0} -\frac{\sin t}{\cos t}\\ &= 0.\end{aligned}$$

We have accomplished something significant. When the derivative $\frac{dy}{dx}$ returns an indeterminate form at $t = t_0$, we can define its value by setting it to be $\lim\limits_{t\to t_0}\frac{dy}{dx}$, if that limit exists. This allows us to find slopes of tangent lines at cusps, which can be very beneficial.

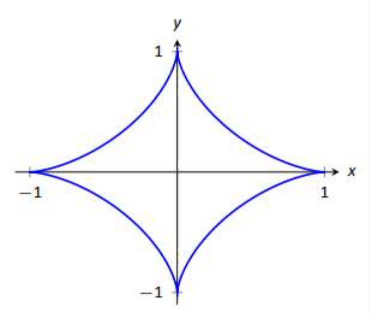


Figure 9.31: *A graph of an astroid.*

We found the slope of the tangent line at $t = 0$ to be 0; therefore the tangent line is $y = 0$, the x- axis.

Concavity

We continue to analyze curves in the plane by considering their concavity; that is, we are interested in $\frac{d^2y}{dx^2}$, "the second derivative of y with respect to x." To find this, we need to find the derivative of $\frac{dy}{dx}$ with respect to x; that is,

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left[\frac{dy}{dx}\right], \tag{9.3.9}$$

but recall that $\frac{dy}{dx}$ is a function of t, not x, making this computation not straightforward.

To make the upcoming notation a bit simpler, let $h(t) = \frac{dy}{dx}$. We want $\frac{d}{dx}[h(t)]$; that is, we want $\frac{dh}{dx}$. We again appeal to the Chain Rule. Note:

$$\frac{dh}{dt} = \frac{dh}{dx} \cdot \frac{dx}{dt} \quad \Rightarrow \quad \frac{dh}{dx} = \frac{dh}{dt} \Big/ \frac{dx}{dt}. \tag{9.3.10}$$

In words, to find $\frac{d^2y}{dx^2}$, we first take the derivative of $\frac{dy}{dx}$ *with respect to* t, then divide by $x'(t)$. We restate this as a Key Idea.

key idea 38 Finding $\frac{d^2y}{dx^2}$ with Parametric Equations

Let $x = f(t)$ and $y = g(t)$ be twice differentiable functions on an open interval I, where $f'(t) \neq 0$ on I. Then

$$\frac{d^2y}{dx^2} \quad = \quad \frac{d}{dt}\left[\frac{dy}{dx}\right] \Big/ \frac{dx}{dt} \quad = \quad \frac{d}{dt}\left[\frac{dy}{dx}\right] \Big/ f'(t). \tag{9.3.11}$$

Examples will help us understand this Key Idea.

Example 9.3.4: Concavity of Plane Curves

Let $x = 5t^2 - 6t + 4$ and $y = t^2 + 6t - 1$ as in Example 9.3.1. Determine the t- intervals on which the graph is concave up/down.

Solution

Concavity is determined by the second derivative of y with respect to x, $\frac{d^2y}{dx^2}$, so we compute that here following Key Idea 38.

In Example 9.3.1, we found $\frac{dy}{dx} = \frac{2t+6}{10t-6}$ and $f'(t) = 10t - 6$. So:

$$\begin{aligned}\frac{d^2y}{dx^2} &= \frac{d}{dt}\left[\frac{2t+6}{10t-6}\right] \Big/ (10t-6) \\ &= -\frac{72}{(10t-6)^2} \Big/ (10t-6) \\ &= -\frac{72}{(10t-6)^3} \\ &= -\frac{9}{(5t-3)^3}\end{aligned}$$

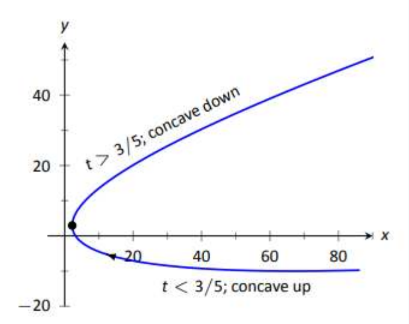
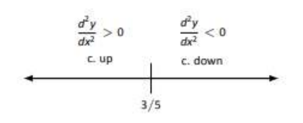


Figure 9.32: Graphing the parametric equations in Example 9.3.4 to demonstrate concavity.

The graph of the parametric functions is concave up when $\frac{d^2y}{dx^2} > 0$ and concave down when $\frac{d^2y}{dx^2} < 0$. We determine the intervals when the second derivative is greater/less than 0 by first finding when it is 0 or undefined.

As the numerator of $-\frac{9}{(5t-3)^3}$ is never 0, $\frac{d^2y}{dx^2} \neq 0$ for all t. It is undefined when $5t - 3 = 0$; that is, when $t = 3/5$. Following the work established in Section 3.4, we look at values of t greater/less than $3/5$ on a number line:



Reviewing Example 9.3.1, we see that when $t = 3/5 = 0.6$, the graph of the parametric equations has a vertical tangent line. This point is also a point of inflection for the graph, illustrated in Figure 9.32.

Example $9.3.5$: Concavity of Plane Curves

Find the points of inflection of the graph of the parametric equations $x = \sqrt{t}$, $y = \sin t$, for $0 \leq t \leq 16$.

Solution

We need to compute $\frac{dy}{dx}$ and $\frac{d^2y}{dx^2}$.

$$\frac{dy}{dx} = \frac{y'(t)}{x'(t)} = \frac{\cos t}{1/(2\sqrt{t})} = 2\sqrt{t}\cos t. \tag{9.3.12}$$

$$\frac{d^2y}{dx^2} = \frac{\frac{d}{dt}\left[\frac{dy}{dx}\right]}{x'(t)} = \frac{\cos t/\sqrt{t} - 2\sqrt{t}\sin t}{1/(2\sqrt{t})} = 2\cos t - 4t\sin t. \tag{9.3.13}$$

The points of inflection are found by setting $\frac{d^2y}{dx^2} = 0$. This is not trivial, as equations that mix polynomials and trigonometric functions generally do not have "nice" solutions.

In Figure 9.33(a) we see a plot of the second derivative. It shows that it has zeros at approximately $t = 0.5,\ 3.5,\ 6.5,\ 9.5,\ 12.5$ and 16. These approximations are not very good, made only by looking at the graph. Newton's Method provides more accurate approximations. Accurate to 2 decimal places, we have:

$$t = 0.65,\ 3.29,\ 6.36,\ 9.48,\ 12.61 \text{ and } 15.74. \tag{9.3.14}$$

The corresponding points have been plotted on the graph of the parametric equations in Figure 9.33(b). Note how most occur near the x- axis, but not exactly on the axis.

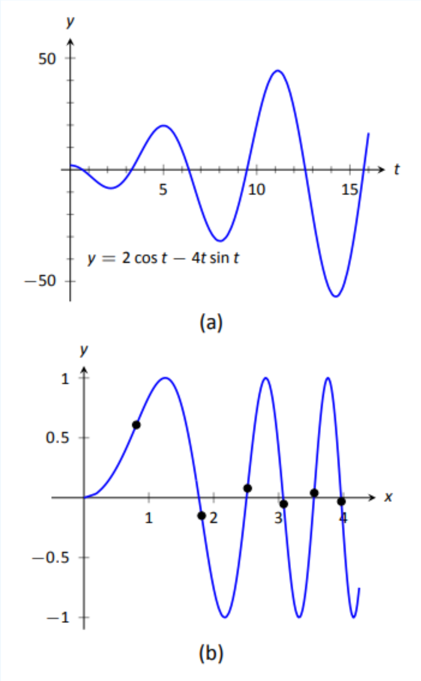


Figure 9.33: In (a), a graph of $\frac{d^2y}{dx^2}$, showing where it is approximately 0. In (b), graph of the parametric equations in Example 9.3.5 along with points of inflection.

Arc Length

We continue our study of the features of the graphs of parametric equations by computing their arc length. Recall in Section 7.4 we found the arc length of the graph of a function, from $x = a$ to $x = b$, to be

$$L = \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx. \tag{9.3.15}$$

We can use this equation and convert it to the parametric equation context. Letting $x = f(t)$ and $y = g(t)$, we know that $\frac{dy}{dx} = g'(t)/f'(t)$. It will also be useful to calculate the differential of x:

$$dx = f'(t)dt \qquad \Rightarrow \qquad dt = \frac{1}{f'(t)} \cdot dx. \tag{9.3.16}$$

Starting with the arc length formula above, consider:

$$\begin{aligned} L &= \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx \\ &= \int_a^b \sqrt{1 + \frac{g'(t)^2}{f'(t)^2}}\, dx. \\ \text{Factor out the } f'(t)^2\text{:} \\ &= \int_a^b \sqrt{f'(t)^2 + g'(t)^2} \cdot \underbrace{\frac{1}{f'(t)}\, dx}_{=dt} \\ &= \int_{t_1}^{t_2} \sqrt{f'(t)^2 + g'(t)^2}\, dt. \end{aligned}$$

Note the new bounds (no longer "x" bounds, but "t" bounds). They are found by finding t_1 and t_2 such that $a = f(t_1)$ and $b = f(t_2)$. This formula is important, so we restate it as a theorem.

theorem 82 arc length of parametric curves

Let $x = f(t)$ and $y = g(t)$ be parametric equations with f' and g' continuous on some open interval I containing t_1 and t_2 on which the graph traces itself only once. The arc length of the graph, from $t = t_1$ to $t = t_2$, is

$$L = \int_{t_1}^{t_2} \sqrt{f'(t)^2 + g'(t)^2}\, dt. \tag{9.3.17}$$

As before, these integrals are often not easy to compute. We start with a simple example, then give another where we approximate the solution.

Example 9.3.6: Arc Length of a Circle

Find the arc length of the circle parametrized by $x = 3\cos t$, $y = 3\sin t$ on $[0, 3\pi/2]$.

Solution
By direct application of Theorem 82, we have

$$\begin{aligned} L &= \int_0^{3\pi/2} \sqrt{(-3\sin t)^2 + (3\cos t)^2}\, dt. \\ \text{Apply the Pythagorean Theorem.} \\ &= \int_0^{3\pi/2} 3\, dt \\ &= 3t\Big|_0^{3\pi/2} = 9\pi/2. \end{aligned}$$

This should make sense; we know from geometry that the circumference of a circle with radius 3 is 6π; since we are finding the arc length of $3/4$ of a circle, the arc length is $3/4 \cdot 6\pi = 9\pi/2$.

Example $9.3.7$: Arc Length of a Parametric Curve

The graph of the parametric equations $x=t(t^2-1)$, $y=t^2-1$ crosses itself as shown in Figure 9.34, forming a "teardrop." Find the arc length of the teardrop.

Solution
We can see by the parametrizations of x and y that when $t=\pm 1$, $x=0$ and $y=0$. This means we'll integrate from $t=-1$ to $t=1$. Applying Theorem 82, we have

$$\begin{aligned} L &= \int_{-1}^{1}\sqrt{(3t^2-1)^2+(2t)^2}\,dt \\ &= \int_{-1}^{1}\sqrt{9t^4-2t^2+1}\,dt. \end{aligned}$$

Unfortunately, the integrand does not have an antiderivative expressible by elementary functions. We turn to numerical integration to approximate its value. Using 4 subintervals, Simpson's Rule approximates the value of the integral as 2.65051. Using a computer, more subintervals are easy to employ, and $n=20$ gives a value of 2.71559. Increasing n shows that this value is stable and a good approximation of the actual value.

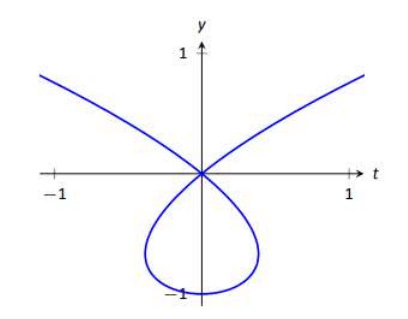


Figure 9.34: A graph of the parametric equations in Example 9.3.7, where the arc length of the teardrop is calculated.

Surface Area of a Solid of Revolution

Related to the formula for finding arc length is the formula for finding surface area. We can adapt the formula found in Key Idea 28 from Section 7.4 in a similar way as done to produce the formula for arc length done before.

key idea 39 Surface Area of a Solid of Revolution

Consider the graph of the parametric equations $x=f(t)$ and $y=g(t)$, where f' and g' are continuous on an open interval I containing t_1 and t_2 on which the graph does not cross itself.

1. The surface area of the solid formed by revolving the graph about the x- axis is (where $g(t)\geq\ 0$ on $[t_1,t_2]$):

$$\text{Surface Area}=2\pi\int_{t_1}^{t_2} g(t)\sqrt{f'(t)^2+g'(t)^2}\,dt. \tag{9.3.18}$$

2. The surface area of the solid formed by revolving the graph about the y- axis is (where $f(t)\geq\ 0$ on $[t_1,t_2]$):

$$\text{Surface Area}=2\pi\int_{t_1}^{t_2} f(t)\sqrt{f'(t)^2+g'(t)^2}\,dt. \tag{9.3.19}$$

Example $9.3.8$: Surface Area of a Solid of Revolution

Consider the teardrop shape formed by the parametric equations $x=t(t^2-1)$, $y=t^2-1$ as seen in Example 9.3.7. Find the surface area if this shape is rotated about the x- axis, as shown in Figure 9.3.8.

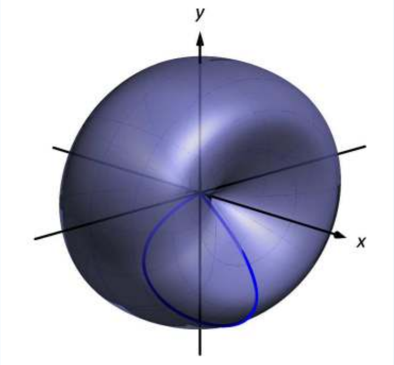


Figure 9.35: Rotating a teardrop shape about the x-axis in Example 9.3.8

Solution

The teardrop shape is formed between $t = -1$ and $t = 1$. Using Key Idea 39, we see we need for $g(t) \geq 0$ on $[-1, 1]$, and this is not the case. To fix this, we simplify replace $g(t)$ with $-g(t)$, which flips the whole graph about the x- axis (and does not change the surface area of the resulting solid). The surface area is:

$$\begin{aligned}\text{Area } S &= 2\pi \int_{-1}^{1} (1-t^2)\sqrt{(3t^2-1)^2+(2t)^2}\ dt \\ &= 2\pi \int_{-1}^{1} (1-t^2)\sqrt{9t^4-2t^2+1}\ dt.\end{aligned}$$

Once again we arrive at an integral that we cannot compute in terms of elementary functions. Using Simpson's Rule with n $= 20$, we find the area to be $S = 9.44$. Using larger values of n shows this is accurate to 2 places after the decimal.

After defining a new way of creating curves in the plane, in this section we have applied calculus techniques to the parametric equation defining these curves to study their properties. In the next section, we define another way of forming curves in the plane. To do so, we create a new coordinate system, called *polar coordinates*, that identifies points in the plane in a manner different than from measuring distances from the y- and x- axes.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 9.3: Calculus and Parametric Equations is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

9.4: Introduction to Polar Coordinates

We are generally introduced to the idea of graphing curves by relating x-values to y-values through a function f. That is, we set $y = f(x)$, and plot lots of point pairs (x, y) to get a good notion of how the curve looks. This method is useful but has limitations, not least of which is that curves that "fail the vertical line test" cannot be graphed without using multiple functions.

The previous two sections introduced and studied a new way of plotting points in the x, y-plane. Using parametric equations, x and y values are computed independently and then plotted together. This method allows us to graph an extraordinary range of curves. This section introduces yet another way to plot points in the plane: using **polar coordinates**.

Polar Coordinates

Start with a point O in the plane called the **pole** (we will always identify this point with the origin). From the pole, draw a ray, called the **initial ray** (we will always draw this ray horizontally, identifying it with the positive x-axis). A point P in the plane is determined by the distance r that P is from O, and the angle θ formed between the initial ray and the segment $\overline{OP}$ (measured counter-clockwise). We record the distance and angle as an ordered pair (r, θ). To avoid confusion with rectangular coordinates, we will denote polar coordinates with the letter P, as in $P(r, \theta)$. This is illustrated in Figure $9.4.1$.

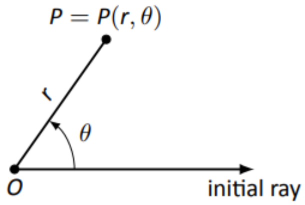


Figure $9.4.1$: Illustrating polar coordinates.

Practice will make this process more clear.

Example $9.4.1$: Plotting Polar Coordinates

Plot the following polar coordinates:

$$A = P(1, \pi/4) \quad B = P(1.5, \pi) \quad C = P(2, -\pi/3) \quad D = P(-1, \pi/4) \tag{9.4.1}$$

Solution

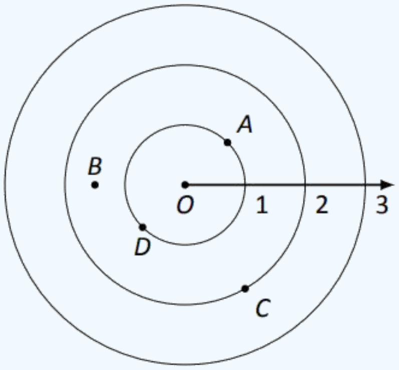


Figure $9.4.2$: Plotting polar points in Example 9.4.1

To aid in the drawing, a polar grid is provided at the bottom of this page. To place the point A, go out 1 unit along the initial ray (putting you on the inner circle shown on the grid), then rotate counter-clockwise $\pi/4$ radians (or 45°). Alternately, one can consider the rotation first: think about the ray from O that forms an angle of $\pi/4$ with the initial ray, then move out 1 unit along this ray (again placing you on the inner circle of the grid).

To plot B, go out 1.5 units along the initial ray and rotate π radians (180°).

To plot C, go out 2 units along the initial ray then rotate *clockwise* $\pi/3$ radians, as the angle given is negative.

> To plot D, move along the initial ray "-1" units -- in other words, "back up" 1 unit, then rotate counter-clockwise by $\pi/4$. The results are given in Figure $9.4.2$.

Consider the following two points: $A = P(1,\pi)$ and $B = P(-1,0)$. To locate A, go out 1 unit on the initial ray then rotate π radians; to locate B, go out -1 units on the initial ray and don't rotate. One should see that A and B are located at the same point in the plane. We can also consider $C = P(1,3\pi)$, or $D = P(1,-\pi)$; all four of these points share the same location.

This ability to identify a point in the plane with multiple polar coordinates is both a "blessing" and a "curse." We will see that it is beneficial as we can plot beautiful functions that intersect themselves (much like we saw with parametric functions). The unfortunate part of this is that it can be difficult to determine when this happens. We'll explore this more later in this section.

Polar to Rectangular Conversion

It is useful to recognize both the rectangular (or, Cartesian) coordinates of a point in the plane and its polar coordinates. Figure $9.4.3$ shows a point P in the plane with rectangular coordinates (x,y) and polar coordinates $P(r,\theta)$. Using trigonometry, we can make the identities given in the following Key Idea.

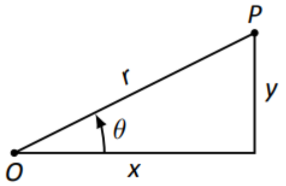


Figure $9.4.3$: Converting between rectangular and polar coordinates.

KEY IDEA 40 Converting Between Rectangular and Polar Coordinates

Given the polar point $P(r,\theta)$, the rectangular coordinates are determined by

$$x = r\cos\theta \qquad y = r\sin\theta. \tag{9.4.2}$$

Given the rectangular coordinates (x,y), the polar coordinates are determined by

$$r^2 = x^2 + y^2 \qquad \tan\theta = \frac{y}{x}. \tag{9.4.3}$$

Example $9.4.2$: Converting Between Polar and Rectangular Coordinates

1. Convert the polar coordinates $P(2,2\pi/3)$ and $P(-1,5\pi/4)$ to rectangular coordinates.
2. Convert the rectangular coordinates $(1,2)$ and $(-1,1)$ to polar coordinates.

1. (a) We start with $P(2,2\pi/3)$. Using Key Idea 40, we have

   $$x = 2\cos(2\pi/3) = -1 \qquad y = 2\sin(2\pi/3) = \sqrt{3}. \tag{9.4.4}$$

   So the rectangular coordinates are $(-1,\sqrt{3}) \approx (-1,1.732)$.

   (b) The polar point $P(-1,5\pi/4)$ is converted to rectangular with:

   $$x = -1\cos(5\pi/4) = \sqrt{2}/2 \qquad y = -1\sin(5\pi/4) = \sqrt{2}/2. \tag{9.4.5}$$

   So the rectangular coordinates are $(\sqrt{2}/2,\sqrt{2}/2) \approx (0.707,0.707)$.

   These points are plotted in Figure $9.4.4$ (a). The rectangular coordinate system is drawn lightly under the polar coordinate system so that the relationship between the two can be seen.
2. (a) To convert the rectangular point $(1,2)$ to polar coordinates, we use the Key Idea to form the following two equations:

   $$1^2 + 2^2 = r^2 \qquad \tan\theta = \frac{2}{1}. \tag{9.4.6}$$

The first equation tells us that $r = \sqrt{5}$. Using the inverse tangent function, we find

$$\tan\theta = 2 \quad \Rightarrow \quad \theta = \tan^{-1} 2 \approx 1.11 \approx 63.43^\circ. \tag{9.4.7}$$

Thus polar coordinates of $(1, 2)$ are $P(\sqrt{5}, 1.11)$.

(b)To convert $(-1, 1)$ to polar coordinates, we form the equations

$$(-1)^2 + 1^2 = r^2 \qquad \tan\theta = \frac{1}{-1}. \tag{9.4.8}$$

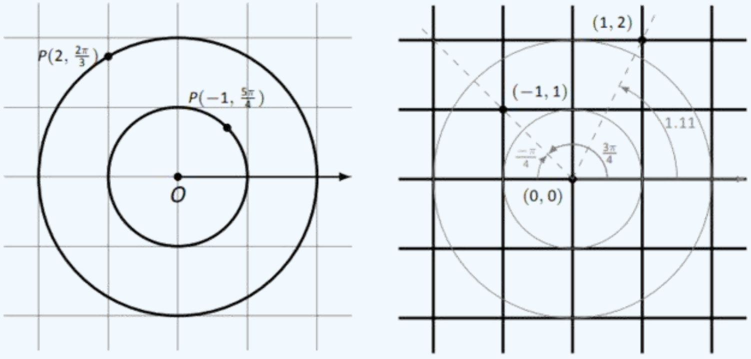
Thus $r = \sqrt{2}$. We need to be careful in computing θ: using the inverse tangent function, we have

$$\tan\theta = -1 \quad \Rightarrow \quad \theta = \tan^{-1}(-1) = -\pi/4 = -45^\circ. \tag{9.4.9}$$

This is not the angle we desire. The range of $\tan^{-1} x$ is $(-\pi/2, \pi/2)$; that is, it returns angles that lie in the 1^{st} and 4^{th} quadrants. To find locations in the 2^{nd} and 3^{rd} quadrants, add π to the result of $\tan^{-1} x$. So $\pi + (-\pi/4)$ puts the angle at $3\pi/4$. Thus the polar point is $P(\sqrt{2}, 3\pi/4)$.

An alternate method is to use the angle θ given by arctangent, but change the sign of r. Thus we could also refer to $(-1, 1)$ as\ $P(-\sqrt{2}, -\pi/4)$.

These points are plotted in Figure $9.4.4$ (b). The polar system is drawn lightly under the rectangular grid with rays to demonstrate the angles used.



(a) (b)

Figure $9.4.4$: Plotting rectangular and polar points in Example 9.4.2

Polar Functions and Polar Graphs

Defining a new coordinate system allows us to create a new kind of function, a polar function. Rectangular coordinates lent themselves well to creating functions that related x and y, such as $y = x^2$. Polar coordinates allow us to create functions that relate r and θ. Normally these functions look like $r = f(\theta)$, although we can create functions of the form $\theta = f(r)$. The following examples introduce us to this concept.

Example $9.4.3$: Introduction to Graphing Polar Functions

Describe the graphs of the following polar functions.

1. $r = 1.5$
2. $\theta = \pi/4$

Solution

1. The equation $r=1.5$ describes all points that are 1.5 units from the pole; as the angle is not specified, any θ is allowable. All points 1.5 units from the pole describes a circle of radius 1.5.

   We can consider the rectangular equivalent of this equation; using $r^2=x^2+y^2$, we see that $1.5^2=x^2+y^2$, which we recognize as the equation of a circle centered at $(0,0)$ with radius 1.5. This is sketched in Figure $9.4.5$.
2. The equation $\theta=\pi/4$ describes all points such that the line through them and the pole make an angle of $\pi/4$ with the initial ray. As the radius r is not specified, it can be any value (even negative). Thus $\theta=\pi/4$ describes the line through the pole that makes an angle of $\pi/4=45^\circ$ with the initial ray.

   We can again consider the rectangular equivalent of this equation. Combine $\tan\theta=y/x$ and $\theta=\pi/4$:

$$\tan\pi/4=y/x \quad \Rightarrow x\tan\pi/4=y \quad \Rightarrow y=x. \tag{9.4.10}$$

   This graph is also plotted in Figure $9.4.5$.

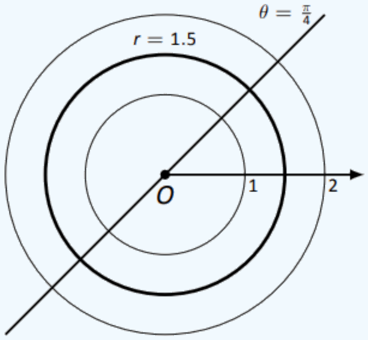


Figure $9.4.5$: Plotting standard polar plots.

The basic rectangular equations of the form $x=h$ and $y=k$ create vertical and horizontal lines, respectively; the basic polar equations $r=h$ and $\theta=\alpha$ create circles and lines through the pole, respectively. With this as a foundation, we can create more complicated polar functions of the form $r=f(\theta)$. The input is an angle; the output is a length, how far in the direction of the angle to go out.

We sketch these functions much like we sketch rectangular and parametric functions: we plot lots of points and "connect the dots" with curves. We demonstrate this in the following example.

Example $9.4.4$: Sketching Polar Functions

Sketch the polar function $r=1+\cos\theta$ on $[0,2\pi]$ by plotting points.

Solution

A common question when sketching curves by plotting points is "Which points should I plot?" With rectangular equations, we often choose "easy" values -- integers, then added more if needed. When plotting polar equations, start with the "common" angles -- multiples of $\pi/6$ and $\pi/4$. Figure $9.4.6$ gives a table of just a few values of θ in $[0,\pi]$.

Consider the point $P(0,2)$ determined by the first line of the table. The angle is 0 radians -- we do not rotate from the initial ray -- then we go out 2 units from the pole. When $\theta=\pi/6$, $r=1.866$ (actually, it is $1+\sqrt{3}/2$); so rotate by $\pi/6$ radians and go out 1.866 units.

The graph shown uses more points, connected with straight lines. (The points on the graph that correspond to points in the table are signified with larger dots.) Such a sketch is likely good enough to give one an idea of what the graph looks like.

θ	$r = 1 + \cos\theta$
0	2
$\pi/6$	1.86603
$\pi/2$	1
$4\pi/3$	0.5
$7\pi/4$	1.70711

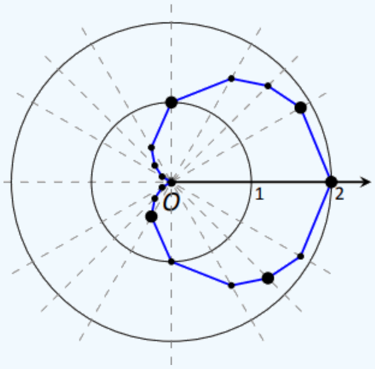


Figure $9.4.6$: Graphing a polar function in Example 9.4.4 by plotting points.

Technology Note

Plotting functions in this way can be tedious, just as it was with rectangular functions. To obtain very accurate graphs, technology is a great aid. Most graphing calculators can plot polar functions; in the menu, set the plotting mode to something like `polar` or `POL`, depending on one's calculator. As with plotting parametric functions, the viewing "window" no longer determines the x-values that are plotted, so additional information needs to be provided. Often with the "window" settings are the settings for the beginning and ending θ values (often called θ_{min} and θ_{max}) as well as the θ_{step} -- that is, how far apart the θ values are spaced. The smaller the θ_{step} value, the more accurate the graph (which also increases plotting time). Using technology, we graphed the polar function $r = 1 + \cos\theta$ from Example 9.4.4 in Figure $9.4.7$.

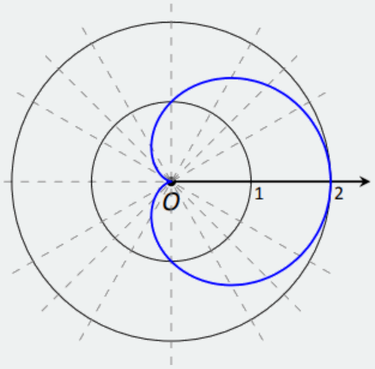


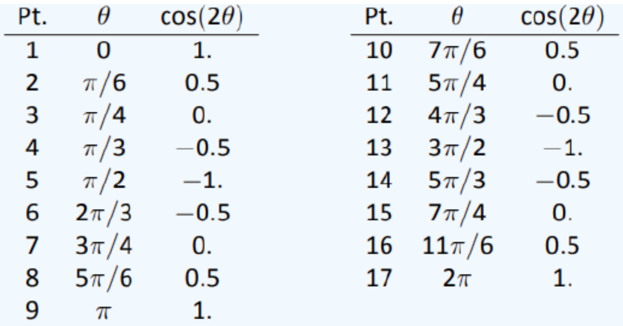
Figure $9.4.7$: Using technology to graph a polar function.

Example $9.4.5$: Sketching Polar Functions

Sketch the polar function $r = \cos(2\theta)$ on $[0, 2\pi]$ by plotting points.

Solution

We start by making a table of $\cos(2\theta)$ evaluated at common angles θ, as shown in Figure $9.4.8$. These points are then plotted in Figure $9.4.9$ (a). This particular graph "moves" around quite a bit and one can easily forget which points should be connected to each other. To help us with this, we numbered each point in the table and on the graph.

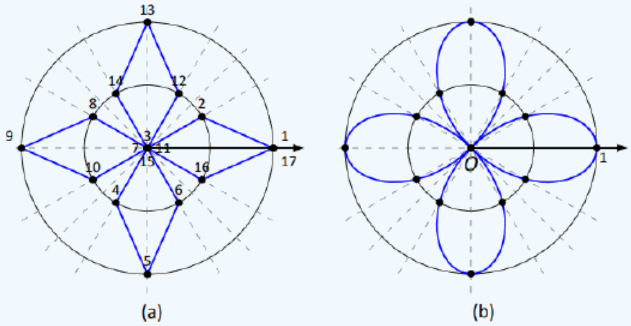


Pt.	θ	$\cos(2\theta)$
1	0	1.
2	$\pi/6$	0.5
3	$\pi/4$	0.
4	$\pi/3$	-0.5
5	$\pi/2$	$-1.$
6	$2\pi/3$	-0.5
7	$3\pi/4$	0.
8	$5\pi/6$	0.5
9	π	1.

Pt.	θ	$\cos(2\theta)$
10	$7\pi/6$	0.5
11	$5\pi/4$	0.
12	$4\pi/3$	-0.5
13	$3\pi/2$	$-1.$
14	$5\pi/3$	-0.5
15	$7\pi/4$	0.
16	$11\pi/6$	0.5
17	2π	1.

Figure 9.4.8: Tables of points for plotting a polar curve.

Using more points (and the aid of technology) a smoother plot can be made as shown in Figure 9.4.9 (b). This plot is an example of a rose curve.



(a)

(b)

Figure 9.4.9: Polar plots from Example 9.4.5

It is sometimes desirable to refer to a graph via a polar equation, and other times by a rectangular equation. Therefore it is necessary to be able to convert between polar and rectangular functions, which we practice in the following example. We will make frequent use of the identities found in Key Idea 40.

Example 9.4.6: Converting between rectangular and polar equations.

Convert from rectangular to polar.

1. $y = x^2$
2. $xy = 1$

Convert from polar to rectangular.

1. $r = \frac{2}{\sin\theta - \cos\theta}$
2. $r = 2\cos\theta$

Solution

1. Replace y with $r\sin\theta$ and replace x with $r\cos\theta$, giving:

$$\begin{aligned} y &= x^2 \\ r\sin\theta &= r^2\cos^2\theta \\ \frac{\sin\theta}{\cos^2\theta} &= r \end{aligned}$$

   We have found that $r = \sin\theta/\cos^2\theta = \tan\theta\sec\theta$. The domain of this polar function is $(-\pi/2, \pi/2)$; plot a few points to see how the familiar parabola is traced out by the polar equation.
2. We again replace x and y using the standard identities and work to solve for r:

$$\begin{aligned} xy &= 1 \\ r\cos\theta \cdot r\sin\theta &= 1 \\ r^2 &= \frac{1}{\cos\theta\sin\theta} \\ r &= \frac{1}{\sqrt{\cos\theta\sin\theta}} \end{aligned}$$

This function is valid only when the product of $\cos\theta\sin\theta$ is positive. This occurs in the first and third quadrants, meaning the domain of this polar function is $(0,\pi/2)\cup(\pi,3\pi/2)$.

We can rewrite the original rectangular equation $xy=1$ as $y=1/x$. This is graphed in Figure $9.4.10$; note how it only exists in the first and third quadrants.

3. There is no set way to convert from polar to rectangular; in general, we look to form the products $r\cos\theta$ and $r\sin\theta$, and then replace these with x and y, respectively. We start in this problem by multiplying both sides by $\sin\theta-\cos\theta$:

$$\begin{aligned} r &= \frac{2}{\sin\theta-\cos\theta} \\ r(\sin\theta-\cos\theta) &= 2 \\ r\sin\theta - r\cos\theta &= 2. \qquad \text{Now replace with } y \text{ and } x\text{:} \\ y-x &= 2 \\ y &= x+2. \end{aligned}$$

The original polar equation, $r=2/(\sin\theta-\cos\theta)$ does not easily reveal that its graph is simply a line. However, our conversion shows that it is. The upcoming gallery of polar curves gives the general equations of lines in polar form.

4. By multiplying both sides by r, we obtain both an r^2 term and an $r\cos\theta$ term, which we replace with x^2+y^2 and x, respectively.

$$\begin{aligned} r &= 2\cos\theta \\ r^2 &= 2r\cos\theta \\ x^2+y^2 &= 2x. \end{aligned}$$

We recognize this as a circle; by completing the square we can find its radius and center.

$$x^2-2x+y^2=0 \tag{9.4.11}$$

$$(x-1)^2+y^2=1. \tag{9.4.12}$$

The circle is centered at $(1,0)$ and has radius 1. The upcoming gallery of polar curves gives the equations of *some* circles in polar form; circles with arbitrary centers have a complicated polar equation that we do not consider here.

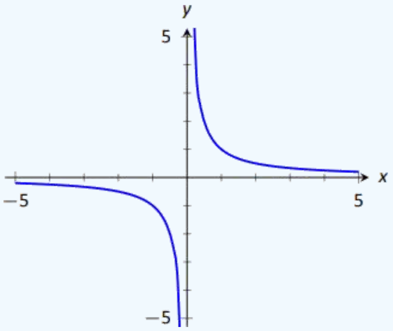


Figure $9.4.10$: Graphing $xy=1$ from Example 9.4.6

Some curves have very simple polar equations but rather complicated rectangular ones. For instance, the equation $r=1+\cos\theta$ describes a *cardiod* (a shape important the sensitivity of microphones, among other things; one is graphed in the gallery in the

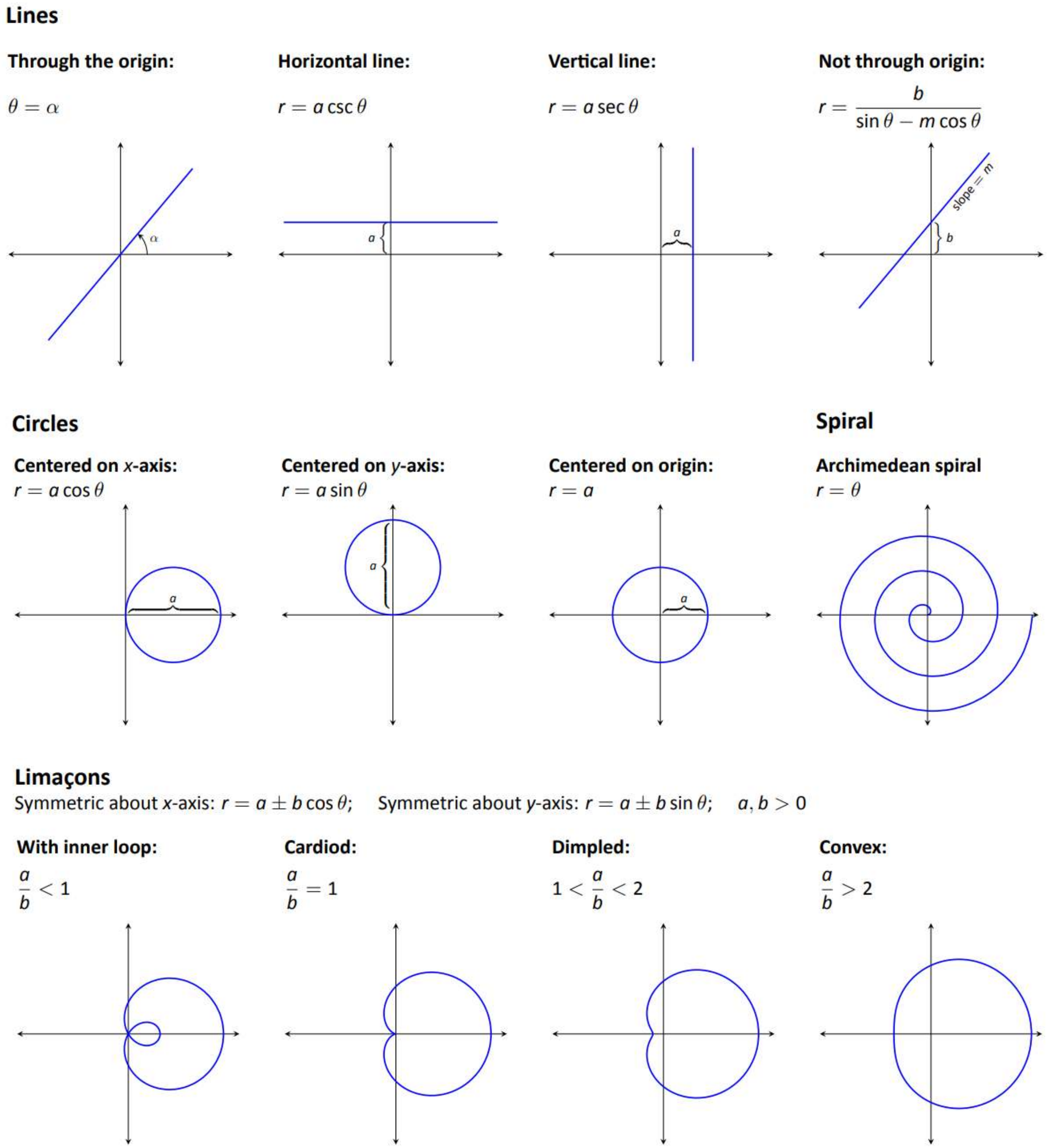
Lima\c con section). It's rectangular form is not nearly as simple; it is the implicit equation

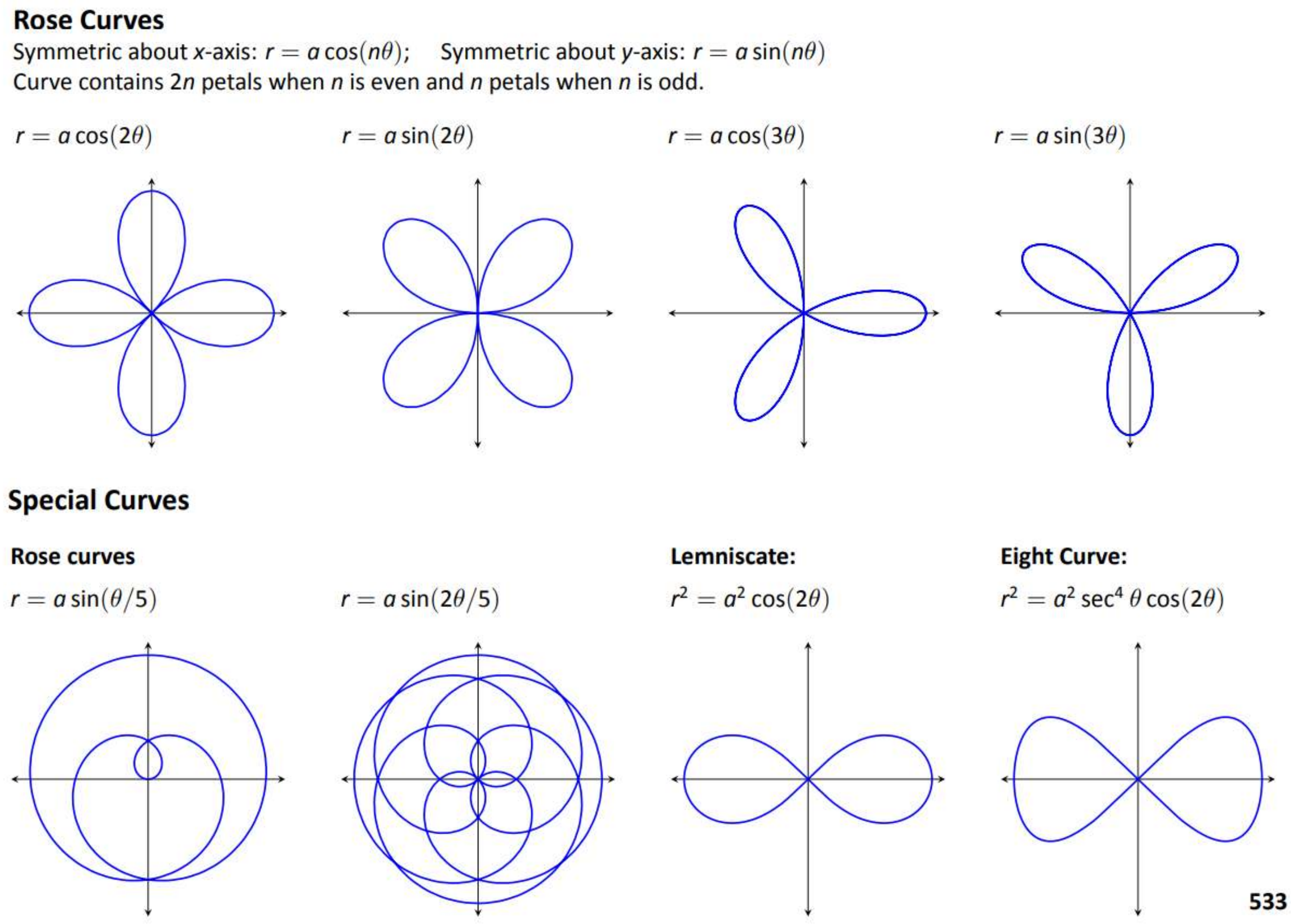
$$x^4+y^4+2x^2y^2-2xy^2-2x^3-y^2=0. \tag{9.4.13}$$

The conversion is not "hard," but takes several steps, and is left as a problem in the Exercise section.

Gallery of Polar Curves

There are a number of basic and "classic" polar curves, famous for their beauty and/or applicability to the sciences. This section ends with a small gallery of some of these graphs. We encourage the reader to understand how these graphs are formed, and to investigate with technology other types of polar functions.


Lines
Through the origin:
$\theta = \alpha$
α
Horizontal line:
$r = a\csc\theta$
a
Vertical line:
$r = a\sec\theta$
a
Not through origin:
$r = \dfrac{b}{\sin\theta - m\cos\theta}$
slope = m
b
Circles
Centered on x-axis:
$r = a\cos\theta$
a
Centered on y-axis:
$r = a\sin\theta$
a
Centered on origin:
$r = a$
a
Spiral
Archimedean spiral
$r = \theta$
Limaçons
Symmetric about x-axis: $r = a \pm b\cos\theta$; Symmetric about y-axis: $r = a \pm b\sin\theta$; $a, b > 0$
With inner loop:
$\frac{a}{b} < 1$
Cardiod:
$\frac{a}{b} = 1$
Dimpled:
$1 < \frac{a}{b} < 2$
Convex:
$\frac{a}{b} > 2$



Rose Curves

Symmetric about x-axis: $r = a\cos(n\theta)$; Symmetric about y-axis: $r = a\sin(n\theta)$

Curve contains $2n$ petals when n is even and n petals when n is odd.

$r = a\cos(2\theta)$ $r = a\sin(2\theta)$ $r = a\cos(3\theta)$ $r = a\sin(3\theta)$

Special Curves

Rose curves

$r = a\sin(\theta/5)$ $r = a\sin(2\theta/5)$

Lemniscate:

$r^2 = a^2\cos(2\theta)$

Eight Curve:

$r^2 = a^2\sec^4\theta\cos(2\theta)$

533

Earlier we discussed how each point in the plane does not have a unique representation in polar form. This can be a "good" thing, as it allows for the beautiful and interesting curves seen in the preceding gallery. However, it can also be a "bad" thing, as it can be difficult to determine where two curves intersect.

Example 9.4.7: Finding points of intersection with polar curves

Determine where the graphs of the polar equations $r = 1 + 3\cos\theta$ and $r = \cos\theta$ intersect.

Solution

As technology is generally readily available, it is usually a good idea to start with a graph. We have graphed the two functions in Figure 9.4.11(a); to better discern the intersection points, part (b) of the figure zooms in around the origin.

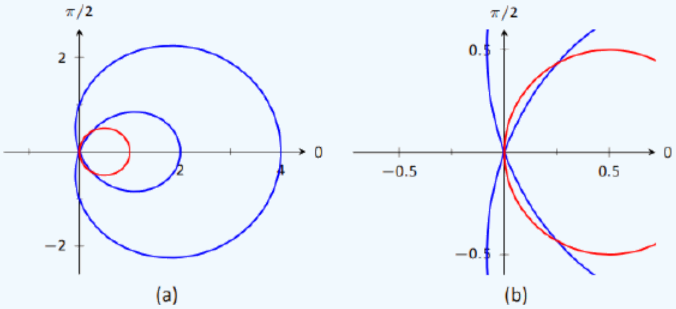


Figure 9.4.11: Graphs to help determine the points of intersection of the polar functions given in Example 9.4.7

We start by setting the two functions equal to each other and solving for θ:

$$\begin{aligned} 1+3\cos\theta &= \cos\theta \\ 2\cos\theta &= -1 \\ \cos\theta &= -\frac{1}{2} \\ \theta &= \frac{2\pi}{3}, \frac{4\pi}{3}. \end{aligned}$$

(There are, of course, infinite solutions to the equation $\cos\theta = -1/2$; as the lima\c con is traced out once on $[0, 2\pi]$, we restrict our solutions to this interval.)

We need to analyze this solution. When $\theta = 2\pi/3$ we obtain the point of intersection that lies in the 4$^{\text{th}}$ quadrant. When $\theta = 4\pi/3$, we get the point of intersection that lies in the 2$^{\text{nd}}$ quadrant. There is more to say about this second intersection point, however. The circle defined by $r = \cos\theta$ is traced out once on $[0, \pi]$, meaning that this point of intersection occurs while tracing out the circle a second time. It seems strange to pass by the point once and then recognize it as a point of intersection only when arriving there a "second time." The first time the circle arrives at this point is when $\theta = \pi/3$.

It is key to understand that these two points are the same: $(\cos\pi/3, \pi/3)$ and $(\cos 4\pi/3, 4\pi/3)$.

To summarize what we have done so far, we have found two points of intersection: when $\theta = 2\pi/3$ and when $\theta = 4\pi/3$. When referencing the circle $r = \cos\theta$, the latter point is better referenced as when $\theta = \pi/3$.

There is yet another point of intersection: the pole (or, the origin). We did not recognize this intersection point using our work above as each graph arrives at the pole at a different θ value.

A graph intersects the pole when $r = 0$. Considering the circle $r = \cos\theta$, $r = 0$ when $\theta = \pi/2$ (and odd multiples thereof, as the circle is repeatedly traced). The lima\c con intersects the pole when $1 + 3\cos\theta = 0$; this occurs when $\cos\theta = -1/3$, or for $\theta = \cos^{-1}(-1/3)$. This is a nonstandard angle, approximately $\theta = 1.9106 = 10$\(9.4.12^\circ\). The lima\c con intersects the pole twice in $[0, 2\pi]$; the other angle at which the lima\c con is at the pole is the reflection of the first angle across the x-axis. That is, $\theta = 4.3726 = 250.53^\circ$.

If all one is concerned with is the (x, y) coordinates at which the graphs intersect, much of the above work is extraneous. We know they intersect at $(0, 0)$; we might not care at what θ value. Likewise, using $\theta = 2\pi/3$ and $\theta = 4\pi/3$ can give us the needed rectangular coordinates. However, in the next section we apply calculus concepts to polar functions. When computing the area of a region bounded by polar curves, understanding the nuances of the points of intersection becomes important.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 9.4: Introduction to Polar Coordinates is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

9.5: Calculus and Polar Functions

The previous section defined polar coordinates, leading to polar functions. We investigated plotting these functions and solving a fundamental question about their graphs, namely, where do two polar graphs intersect?

We now turn our attention to answering other questions, whose solutions require the use of calculus. A basis for much of what is done in this section is the ability to turn a polar function $r = f(\theta)$ into a set of parametric equations. Using the identities $x = r\cos\theta$ and $y = r\sin\theta$, we can create the parametric equations $x = f(\theta)\cos\theta$, $y = f(\theta)\sin\theta$ and apply the concepts of Section 9.3.

Polar Functions and $\frac{dy}{dx}$

We are interested in the lines tangent a given graph, regardless of whether that graph is produced by rectangular, parametric, or polar equations. In each of these contexts, the slope of the tangent line is $\frac{dy}{dx}$. Given $r = f(\theta)$, we are generally *not* concerned with $r' = f'(\theta)$; that describes how fast r changes with respect to θ. Instead, we will use $x = f(\theta)\cos\theta$, $y = f(\theta)\sin\theta$ to compute $\frac{dy}{dx}$.

Using Key Idea 37 we have

$$\frac{dy}{dx} = \frac{dy}{d\theta} \Big/ \frac{dx}{d\theta}. \tag{9.5.1}$$

Each of the two derivatives on the right hand side of the equality requires the use of the Product Rule. We state the important result as a Key Idea.

key idea 41 Finding $\frac{dy}{dx}$ with Polar Functions

Let $r = f(\theta)$ be a polar function. With $x = f(\theta)\cos\theta$ and $y = f(\theta)\sin\theta$,

$$\frac{dy}{dx} = \frac{f'(\theta)\sin\theta + f(\theta)\cos\theta}{f'(\theta)\cos\theta - f(\theta)\sin\theta}. \tag{9.5.2}$$

Example 9.5.1: Finding $\frac{dy}{dx}$ with polar functions.

Consider the limacon $r = 1 + 2\sin\theta$ on $[0, 2\pi]$.

1. Find the equations of the tangent and normal lines to the graph at $\theta = \pi/4$.
2. Find where the graph has vertical and horizontal tangent lines.

Solution

1. We start by computing $\frac{dy}{dx}$. With $f'(\theta) = 2\cos\theta$, we have

$$\begin{aligned} \frac{dy}{dx} &= \frac{2\cos\theta\sin\theta + \cos\theta(1+2\sin\theta)}{2\cos^2\theta - \sin\theta(1+2\sin\theta)} \\ &= \frac{\cos\theta(4\sin\theta + 1)}{2(\cos^2\theta - \sin^2\theta) - \sin\theta}. \end{aligned}$$

When $\theta = \pi/4$, $\frac{dy}{dx} = -2\sqrt{2} - 1$ (this requires a bit of simplification). In rectangular coordinates, the point on the graph at $\theta = \pi/4$ is $(1 + \sqrt{2}/2, 1 + \sqrt{2}/2)$. Thus the rectangular equation of the line tangent to the limacon at $\theta = \pi/4$ is

$$y = (-2\sqrt{2} - 1)\Big(x - (1 + \sqrt{2}/2)\Big) + 1 + \sqrt{2}/2 \approx -3.83x + 8.24. \tag{9.5.3}$$

The limacon and the tangent line are graphed in Figure 9.47.
The normal line has the opposite--reciprocal slope as the tangent line, so its equation is

$$y \approx \frac{1}{3.83}x + 1.26. \tag{9.5.4}$$

2. To find the horizontal lines of tangency, we find where $\frac{dy}{dx} = 0$; thus we find where the numerator of our equation for $\frac{dy}{dx}$ is 0.

$$\cos\theta(4\sin\theta+1)=0 \quad\Rightarrow\quad \cos\theta=0 \quad\text{or}\quad 4\sin\theta+1=0. \tag{9.5.5}$$

On $[0,2\pi]$, $\cos\theta=0$ when $\theta=\pi/2,\ 3\pi/2$.

Setting $4\sin\theta+1=0$ gives $\theta=\sin^{-1}(-1/4)\approx-0.2527=-14.48^\circ$. We want the results in $[0,2\pi]$; we also recognize there are two solutions, one in the 3^{rd} quadrant and one in the 4^{th}. Using reference angles, we have our two solutions as $\theta=3.39$ and 6.03 radians. The four points we obtained where the limacon has a horizontal tangent line are given in Figure 9.47 with black--filled dots.

To find the vertical lines of tangency, we set the denominator of $\frac{dy}{dx}=0$.

$$\begin{gathered}
2(\cos^2\theta-\sin^2\theta)-\sin\theta=0.\\
\text{Convert the } \cos^2\theta \text{ term to } 1-\sin^2\theta\text{:}\\
2(1-\sin^2\theta-\sin^2\theta)-\sin\theta=0\\
4\sin^2\theta+\sin\theta-1=0.\\
\text{Recognize this as a quadratic in the variable } \sin\theta\text{. Using the quadratic formula, we have}\\
\sin\theta=\frac{-1\pm\sqrt{33}}{8}.
\end{gathered}$$

We solve $\sin\theta=\frac{-1+\sqrt{33}}{8}$ and $\sin\theta=\frac{-1-\sqrt{33}}{8}$:

$$\begin{aligned}
\sin\theta&=\frac{-1+\sqrt{33}}{8} & \sin\theta&=\frac{-1-\sqrt{33}}{8}\\
\theta&=\sin^{-1}\left(\frac{-1+\sqrt{33}}{8}\right) & \theta&=\sin^{-1}\left(\frac{-1-\sqrt{33}}{8}\right)\\
\theta&=0.6399 & \theta&=-1.0030
\end{aligned}$$

In each of the solutions above, we only get one of the possible two solutions as $\sin^{-1}x$ only returns solutions in $[-\pi/2,\pi/2]$, the 4^{th} and 1^{st} quadrants. Again using reference angles, we have:

$$\sin\theta=\frac{-1+\sqrt{33}}{8} \quad\Rightarrow\quad \theta=0.6399,\ 3.7815 \text{ radians} \tag{9.5.6}$$

and

$$\sin\theta=\frac{-1-\sqrt{33}}{8} \quad\Rightarrow\quad \theta=4.1446,\ 5.2802 \text{ radians.} \tag{9.5.7}$$

These points are also shown in Figure 9.47 with white--filled dots.

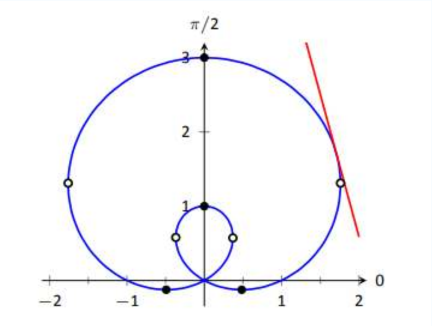


Figure 9.47: The limacon in Example 9.5.1 with its tangent line at $\theta=\pi/4$ and points of vertical and horizontal tangency.

When the graph of the polar function $r=f(\theta)$ intersects the pole, it means that $f(\alpha)=0$ for some angle α. Thus the formula for $\frac{dy}{dx}$ in such instances is very simple, reducing simply to

$$\frac{dy}{dx}=\tan\alpha. \tag{9.5.8}$$

This equation makes an interesting point. It tells us the slope of the tangent line at the pole is $\tan\alpha$; some of our previous work (see, for instance, Example 9.4.3) shows us that the line through the pole with slope $\tan\alpha$ has polar equation $\theta=\alpha$. Thus when a polar graph touches the pole at $\theta=\alpha$, the equation of the tangent line at the pole is $\theta=\alpha$.

Example $9.5.2$: Finding tangent lines at the pole.

Let $r=1+2\sin\theta$, a limacon. Find the equations of the lines tangent to the graph at the pole.

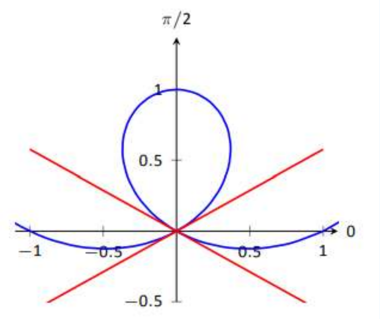


Figure 9.48: Graphing the tangent lines at the pole in Example 9.5.2.

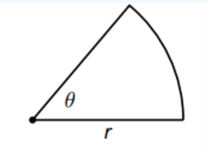
Solution

We need to know when $r=0$.

$$\begin{aligned}1+2\sin\theta&=0\\ \sin\theta&=-1/2\\ \theta&=\frac{7\pi}{6},\ \frac{11\pi}{6}.\end{aligned}$$

Thus the equations of the tangent lines, in polar, are $\theta=7\pi/6$ and $\theta=11\pi/6$. In rectangular form, the tangent lines are $y=\tan(7\pi/6)x$ and $y=\tan(11\pi/6)x$. The full limacon con can be seen in Figure 9.47; we zoom in on the tangent lines in Figure 9.48.

Note: Recall that the area of a sector of a circle with radius r subtended by an angle θ is $A=\frac{1}{2}\theta r^2$.



Area

When using rectangular coordinates, the equations $x=h$ and $y=k$ defined vertical and horizontal lines, respectively, and combinations of these lines create rectangles (hence the name "rectangular coordinates"). It is then somewhat natural to use rectangles to approximate area as we did when learning about the definite integral.

When using polar coordinates, the equations $\theta=\alpha$ and $r=c$ form lines through the origin and circles centered at the origin, respectively, and combinations of these curves form sectors of circles. It is then somewhat natural to calculate the area of regions defined by polar functions by first approximating with sectors of circles.

Consider Figure 9.49 (a) where a region defined by $r=f(\theta)$ on $[\alpha,\beta]$ is given. (Note how the "sides" of the region are the lines $\theta=\alpha$ and $\theta=\beta$, whereas in rectangular coordinates the "sides" of regions were often the vertical lines $x=a$ and $x=b$.)

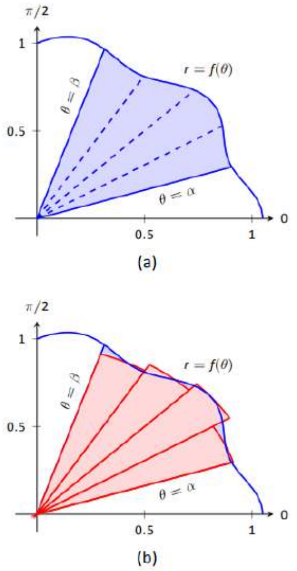


Figure 9.49: Computing the area of a polar region.

Partition the interval $[\alpha, \beta]$ into n equally spaced subintervals as $\alpha = \theta_1 < \theta_2 < \cdots < \theta_{n+1} = \beta$. The length of each subinterval is $\Delta\theta = (\beta - \alpha)/n$, representing a small change in angle. The area of the region defined by the i^{th} subinterval $[\theta_i, \theta_{i+1}]$ can be approximated with a sector of a circle with radius $f(c_i)$, for some c_i in $[\theta_i, \theta_{i+1}]$. The area of this sector is $\frac{1}{2} f(c_i)^2 \Delta\theta$. This is shown in part (b) of the figure, where $[\alpha, \beta]$ has been divided into 4 subintervals. We approximate the area of the whole region by summing the areas of all sectors:

$$\text{Area} \approx \sum_{i=1}^{n} \frac{1}{2} f(c_i)^2 \Delta\theta. \tag{9.5.9}$$

This is a Riemann sum. By taking the limit of the sum as $n \to \infty$, we find the exact area of the region in the form of a definite integral.

THEOREM 83 AREA OF A POLAR REGION

Let f be continuous and non-negative on $[\alpha, \beta]$, where $0 \le \beta - \alpha \le 2\pi$. The area A of the region bounded by the curve $r = f(\theta)$ and the lines $\theta = \alpha$ and $\theta = \beta$ is

$$A = \frac{1}{2} \int_\alpha^\beta f(\theta)^2 \, d\theta = \frac{1}{2} \int_\alpha^\beta r^2 \, d\theta \tag{9.5.10}$$

The theorem states that $0 \le \beta - \alpha \le 2\pi$. This ensures that region does not overlap itself, which would give a result that does not correspond directly to the area.

Example $9.5.3$: Area of a polar region

Find the area of the circle defined by $r = \cos\theta$. (Recall this circle has radius $1/2$.)

Solution

This is a direct application of Theorem 83. The circle is traced out on $[0, \pi]$, leading to the integral

$$\begin{aligned} \text{Area} &= \frac{1}{2} \int_0^\pi \cos^2\theta \, d\theta \\ &= \frac{1}{2} \int_0^\pi \frac{1 + \cos(2\theta)}{2} \, d\theta \\ &= \frac{1}{4} \left(\theta + \frac{1}{2} \sin(2\theta) \right) \Big|_0^\pi \\ &= \frac{1}{4} \pi. \end{aligned}$$

Of course, we already knew the area of a circle with radius $1/2$. We did this example to demonstrate that the area formula is correct.

Note: Example 9.5.3 requires the use of the integral $\int \cos^2\theta\ d\theta$. This is handled well by using the power reducing formula as found in the back of this text. Due to the nature of the area formula, integrating $\cos^2\theta$ and $\sin^2\theta$ is required often. We offer here these indefinite integrals as a time--saving measure.

$$\int \cos^2\theta\ d\theta = \frac{1}{2}\theta + \frac{1}{4}\sin(2\theta) + C \tag{9.5.11}$$

$$\int \sin^2\theta\ d\theta = \frac{1}{2}\theta - \frac{1}{4}\sin(2\theta) + C \tag{9.5.12}$$

Example $9.5.4$: Area of a polar region

Find the area of the cardiod $r = 1+\cos\theta$ bound between $\theta = \pi/6$ and $\theta = \pi/3$, as shown in Figure 9.50.

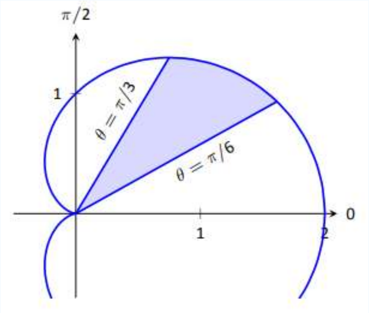


Figure 9.50: Finding the area of the shaded region of a cardiod in Example 9.5.4

Solution

This is again a direct application of Theorem 83.

$$\begin{aligned} \text{Area} &= \frac{1}{2}\int_{\pi/6}^{\pi/3} (1+\cos\theta)^2\ d\theta \\ &= \frac{1}{2}\int_{\pi/6}^{\pi/3} (1+2\cos\theta+\cos^2\theta)\ d\theta \\ &= \frac{1}{2}\left(\theta + 2\sin\theta + \frac{1}{2}\theta + \frac{1}{4}\sin(2\theta)\right)\Bigg|_{\pi/6}^{\pi/3} \\ &= \frac{1}{8}\left(\pi + 4\sqrt{3} - 4\right) \approx 0.7587. \end{aligned}$$

Area Between Curves

Our study of area in the context of rectangular functions led naturally to finding area bounded between curves. We consider the same in the context of polar functions. \index{polar!functions!area between curves}

Consider the shaded region shown in Figure 9.51. We can find the area of this region by computing the area bounded by $r_2 = f_2(\theta)$ and subtracting the area bounded by $r_1 = f_1(\theta)$ on $[\alpha, \beta]$. Thus

$$\text{Area} \ = \ \frac{1}{2}\int_\alpha^\beta r_2^2\ d\theta - \frac{1}{2}\int_\alpha^\beta r_1^2\ d\theta = \frac{1}{2}\int_\alpha^\beta \left(r_2^2 - r_1^2\right)\ d\theta. \tag{9.5.13}$$

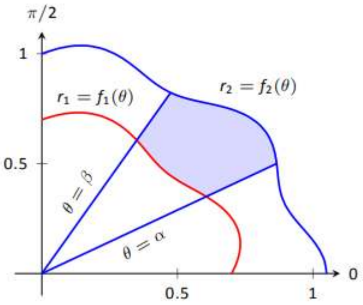


Figure 9.51: Illustrating area bound between two polar curves.

KEY IDEA 42 area between polar curves

The area A of the region bounded by $r_1 = f_1(\theta)$ and $r_2 = f_2(\theta)$, $\theta = \alpha$ and $\theta = \beta$, where $f_1(\theta) \le f_2(\theta)$ on $[\alpha, \beta]$, is

$$A = \frac{1}{2}\int_{\alpha}^{\beta} \left(r_2^2 - r_1^2\right)\, d\theta. \tag{9.5.14}$$

Example $9.5.5$: Area between polar curves

Find the area bounded between the curves $r = 1 + \cos\theta$ and $r = 3\cos\theta$, as shown in Figure 9.52.

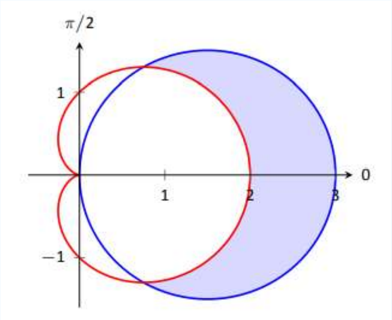


Figure 9.52: Finding the area between polar curves in Example 9.5.5.

Solution

We need to find the points of intersection between these two functions. Setting them equal to each other, we find:

$$\begin{aligned} 1 + \cos\theta &= 3\cos\theta \\ \cos\theta &= 1/2 \\ \theta &= \pm\pi/3 \end{aligned}$$

Thus we integrate $\frac{1}{2}\left((3\cos\theta)^2 - (1+\cos\theta)^2\right)$ on $[-\pi/3, \pi/3]$.

$$\begin{aligned} \text{Area} &= \frac{1}{2}\int_{-\pi/3}^{\pi/3} \left((3\cos\theta)^2 - (1+\cos\theta)^2\right)\, d\theta \\ &= \frac{1}{2}\int_{-\pi/3}^{\pi/3} \left(8\cos^2\theta - 2\cos\theta - 1\right)\, d\theta \\ &= \left(2\sin(2\theta) - 2\sin\theta + 3\theta\right)\Big|_{-\pi/3}^{\pi/3} \\ &= 2\pi. \end{aligned}$$

Amazingly enough, the area between these curves has a "nice" value

Example $9.5.6$: Area defined by polar curves

Find the area bounded between the polar curves $r = 1$ and $r = 2\cos(2\theta)$, as shown in Figure 9.53 (a).

Solution

We need to find the point of intersection between the two curves. Setting the two functions equal to each other, we have

$$2\cos(2\theta)=1 \quad \Rightarrow \quad \cos(2\theta)=\frac{1}{2} \quad \Rightarrow \quad 2\theta=\pi/3 \quad \Rightarrow \quad \theta=\pi/6. \tag{9.5.15}$$

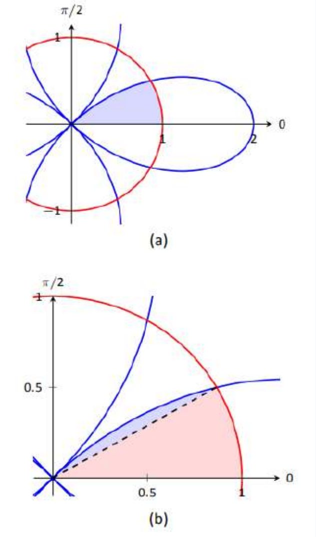


Figure 9.53: Graphing the region bounded by the functions in Example 9.5.6

In part (b) of the figure, we zoom in on the region and note that it is not really bounded *between* two polar curves, but rather *by* two polar curves, along with $\theta=0$. The dashed line breaks the region into its component parts. Below the dashed line, the region is defined by $r=1$, $\theta=0$ and $\theta=\pi/6$. (Note: the dashed line lies on the line $\theta=\pi/6$.) Above the dashed line the region is bounded by $r=2\cos(2\theta)$ and $\theta=\pi/6$. Since we have two separate regions, we find the area using two separate integrals.

Call the area below the dashed line A_1 and the area above the dashed line A_2. They are determined by the following integrals:

$$A_1=\frac{1}{2}\int_0^{\pi/6}(1)^2\,d\theta \qquad A_2=\frac{1}{2}\int_{\pi/6}^{\pi/4}\big(2\cos(2\theta)\big)^2\,d\theta. \tag{9.5.16}$$

(The upper bound of the integral computing A_2 is $\pi/4$ as $r=2\cos(2\theta)$ is at the pole when $\theta=\pi/4$.)

We omit the integration details and let the reader verify that $A_1=\pi/12$ and $A_2=\pi/12-\sqrt{3}/8$; the total area is $A=\pi/6-\sqrt{3}/8$.

Arc Length

As we have already considered the arc length of curves defined by rectangular and parametric equations, we now consider it in the context of polar equations. Recall that the arc length L of the graph defined by the parametric equations $x=f(t)$, $y=g(t)$ on $[a,b]$ is

$$L=\int_a^b\sqrt{f'(t)^2+g'(t)^2}\,dt=\int_a^b\sqrt{x'(t)^2+y'(t)^2}\,dt. \tag{9.5.17}$$

Now consider the polar function $r=f(\theta)$. We again use the identities $x=f(\theta)\cos\theta$ and $y=f(\theta)\sin\theta$ to create parametric equations based on the polar function. We compute $x'(\theta)$ and $y'(\theta)$ as done before when computing $\frac{dy}{dx}$, then apply Equation 9.5.17.

The expression $x'(\theta)^2+y'(\theta)^2$ can be simplified a great deal; we leave this as an exercise and state that

$$x'(\theta)^2+y'(\theta)^2=f'(\theta)^2+f(\theta)^2. \tag{9.5.18}$$

This leads us to the arc length formula.

key idea 43 arc length of polar curves

Let $r = f(\theta)$ be a polar function with f' continuous on an open interval I containing $[\alpha, \beta]$, on which the graph traces itself only once. The arc length L of the graph on $[\alpha, \beta]$ is

$$L = \int_\alpha^\beta \sqrt{f'(\theta)^2 + f(\theta)^2}\, d\theta = \int_\alpha^\beta \sqrt{(r')^2 + r^2}\, d\theta. \tag{9.5.19}$$

Example 9.5.7: Arc length of a limacon

Find the arc length of the limacon $r = 1 + 2\sin t$.

Solution

With $r = 1 + 2\sin t$, we have $r' = 2\cos t$. The limacon is traced out once on $[0, 2\pi]$, giving us our bounds of integration. Applying Key Idea 43, we have

$$\begin{aligned} L &= \int_0^{2\pi} \sqrt{(2\cos\theta)^2 + (1 + 2\sin\theta)^2}\, d\theta \\ &= \int_0^{2\pi} \sqrt{4\cos^2\theta + 4\sin^2\theta + 4\sin\theta + 1}\, d\theta \\ &= \int_0^{2\pi} \sqrt{4\sin\theta + 5}\, d\theta \\ &\approx 13.3649. \end{aligned}$$

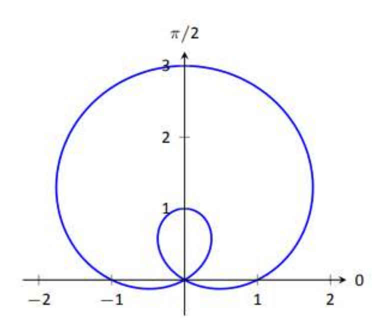


Figure 9.54: The limacon in Example 9.5.7 whose arc length is measured.

The final integral cannot be solved in terms of elementary functions, so we resorted to a numerical approximation. (Simpson's Rule, with $n = 4$, approximates the value with 13.0608. Using $n = 22$ gives the value above, which is accurate to 4 places after the decimal.)

Surface Area

The formula for arc length leads us to a formula for surface area. The following Key Idea is based on Key Idea 39.

KEY IDEA 44 SURFACE AREA OF A SOLID OF REVOLUTION

Consider the graph of the polar equation $r = f(\theta)$, where f' is continuous on an open interval containing $[\alpha, \beta]$ on which the graph does not cross itself.

1. The surface area of the solid formed by revolving the graph about the initial ray ($\theta = 0$) is:

$$\text{Surface Area} = 2\pi \int_\alpha^\beta f(\theta)\sin\theta \sqrt{f'(\theta)^2 + f(\theta)^2}\, d\theta. \tag{9.5.20}$$

2. The surface area of the solid formed by revolving the graph about the line $\theta = \pi/2$ is:

$$\text{Surface Area} = 2\pi \int_\alpha^\beta f(\theta)\cos\theta \sqrt{f'(\theta)^2 + f(\theta)^2}\, d\theta. \tag{9.5.21}$$

Example 9.5.8: Surface area determined by a polar curve

Find the surface area formed by revolving one petal of the rose curve $r = \cos(2\theta)$ about its central axis (see Figure 9.55.

Solution

We choose, as implied by the figure, to revolve the portion of the curve that lies on $[0, \pi/4]$ about the initial ray. Using Key Idea ??? and the fact that $f'(\theta) = -2\sin(2\theta)$, we have

$$\begin{aligned}\text{Surface Area} &= 2\pi \int_0^{\pi/4} \cos(2\theta)\sin(\theta)\sqrt{\big(-2\sin(2\theta)\big)^2 + \big(\cos(2\theta)\big)^2}\, d\theta \\ &\approx 1.36707.\end{aligned}$$

The integral is another that cannot be evaluated in terms of elementary functions. Simpson's Rule, with $n = 4$, approximates the value at $1.36751.$%; with $n = 10$, the value is accurate to 4 decimal places.

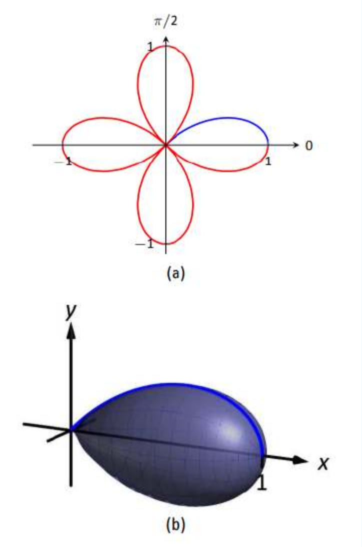


Figure 9.55: Finding the surface area of a rose-curve petal that is revolved around its central axis.

This chapter has been about curves in the plane. While there is great mathematics to be discovered in the two dimensions of a plane, we live in a three dimensional world and hence we should also look to do mathematics in 3D -- that is, in *space*. The next chapter begins our exploration into space by introducing the topic of *vectors*, which are incredibly useful and powerful mathematical objects.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 9.5: Calculus and Polar Functions is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

9.E: Applications of Curves in a Plane (Exercises)

9.1: Conic Sections

Terms and Concepts

1. What is the difference between degenerate and nondegenerate conics?

2. Use your own words to explain what the eccentricity of an ellipse measures.

3. What has the largest eccentricity: an ellipse or a hyperbola?

4. Explain why the following is true: “If the coefficient of the x^2 term in the equation of an ellipse in standard form is smaller than the coefficient of the y^2 term, then the ellipse has a horizontal major axis.”

5. Explain how one can quickly look at the equation of a hyperbola in standard form and determine whether the transverse axis is horizontal or vertical.

Problems

In Exercises 6-13, find the equation of the parabola defined by the given information. Sketch the parabola.

6. Focus: $(3,2)$; directrix: $y=1$

7. Focus: $(-1,-4)$; directrix: $y=2$

8. Focus: $(1,5)$; directrix: $x=3$

9. Focus: $(1/4,0)$; directrix: $x=-1/4$

10. Focus: $(1,1)$; vertex: $(1,2)$

11. Focus: $(-3,0)$; vertex: $(0,0)$

12. Vertex: $(0,0)$; directrix: $y=-1/16$

13. Vertex: $(2,3)$; directrix: $x=4$

In Exercises 14-15, the equation of a parabola and a point on its graph are given. Find the focus and directrix of the parabola, and verify that the given point is equidistant from the focus and directrix.

14. $y=\frac{1}{4}x^2$, $P=(2,1)$

15. $x=\frac{1}{8}(y-2)^2+3,\ P=(11,10)$

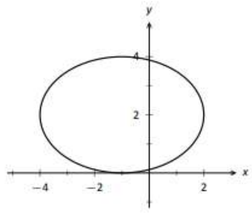
In Exercises 16-17, sketch the ellipse defined by the given equation. Label the center, foci and vertices.

16. $\frac{(x-1)^2}{3}+\frac{(y-2)^2}{5}=1$

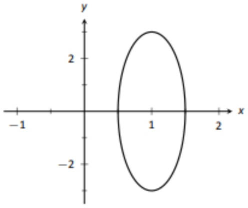
17. $\frac{1}{25}x^2+\frac{1}{9}(y+3)^2=1$

In Exercises 18-19, find the equation of the ellipse shown in the graph. Give the location of the foci and the eccentricity of the ellipse.

18.



19.



In Exercises 20-23, find the equation of the ellipse defined by the given information. Sketch the ellipse.

20. Foci: $(\pm 2, 0)$; vertices: $(\pm 3, 0)$

21. Foci: $(-1, 3) \text{ and } (5, 3)$; vertices: $(-3, 3) \text{ and } (7, 3)$

22. Foci: $(2, \pm 2)$; vertices: $(2, \pm 7)$

23. Focus: $(-1, 5)$; vertex: $(-1, -4)$; center: $(-1, 1)$

In Exercises 24-27, write the equation of the given ellipse in standard form.

24. $x^2 - 2x + 2y^2 - 8y = -7$

25. $5x^2 + 3y^2 = 15$

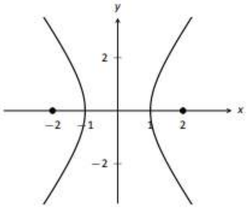
26. $3x^2 + 2y^2 - 12y + 6 = 0$

27. $x^2 + y^2 - 4x - 4y + 4 = 0$

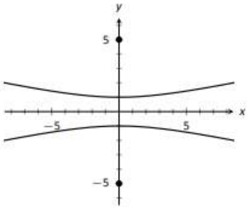
28. Consider the ellipse given by $\frac{(x-1)^2}{4} + \frac{(y-3)^2}{12} = 1$.
(a) Verify that the foci are located at $(1, 3 \pm 2\sqrt{2})$
(b) The point $P_1 = (2, 6) \text{ and } P_2 = (1 + \sqrt{2}, 3 + \sqrt{6}) \approx (2.414, 5.449)$ lie on the ellipse. Verify that the sum of distances from each point to the foci is the same.

In Exercises 29-32, find the equation of the hyperbola shown in the graph.

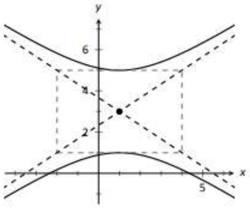
29.



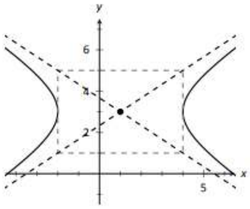
30.



31.



32.



In Exercises 33-34, sketch the hyperbola defined by the given equation. Label the center and foci.

33. $\frac{(x-1)^2}{16} - \frac{(y+2)^2}{9} = 1$

34. $(y-4)^2 - \frac{(x+1)^2}{25} = 1$

In Exercises 35-38, find the equation of the hyperbola defined by the given information. Sketch the hyperbola.

35. Foci: $(\pm 3, 0)$; vertices: $(\pm 2, 0)$

36. Foci: $(0, \pm 3)$; vertices: $(0, \pm 2)$

37. Foci: $(-2, 3)$ and $(8, 3)$ vertices: $(-1, 3)$ and $(7, 3)$

38. Foci: $(3, -2)$ and $(3, 8)$ vertices: $(3, 0)$ and $(3, 6)$

In Exercises 39-42, write the equation of the hyperbola in standard form.

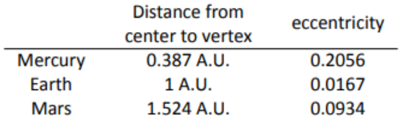
39. $3x^2 - 4y^2 = 12$

40. $3x^2 - y^2 + 2y = 10$

41. $x^2 - 10y^2 + 40y = 30$

42. $(4y - x)(4y + x) = 4$

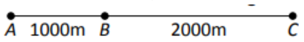
43. Johannes Kepler discovered that the planets of our solar system have elliptical orbits with the Sun at one focus. The Earth's elliptical orbit is used as a standard unit of distance; the distance from the center of Earth's elliptical orbit to one vertex is 1 Astronomical Unit, or A.U.
The following table gives information about the orbits of three planets.



	Distance from center to vertex	eccentricity
Mercury	0.387 A.U.	0.2056
Earth	1 A.U.	0.0167
Mars	1.524 A.U.	0.0934

(a) In an ellipse, knowing $c^2 = a^2 - b^2$ and $e = c/a$ allows us to find b in terms of a and e. Show $b = a\sqrt{1-e^2}$.
(b) For each planet, find equations of their elliptical orbit of the form $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$. (This places the center at (0,0), but the Sun is in a different location for each planet.)
(c) Shift the equations so that the Sun lies at the origin. Plot the three elliptical orbits.

44. A loud sound is recorded at three stations that lie on a line as shown in the figure below. Station A recorded the sound 1 second after Station B, and Station C recorded the sound 3 seconds after B. Using the speed of sound as 340m/s, determine the location of the sound's origination.



9.2: Parametric Equations

Terms and Concepts

1. T/F: When sketching the graph of parametric equations, the x and y values are found separately, then plotted together.

2. The direction in which a graph is "moving" is called the _____ of the graph.

3. An Equation written as $y = f(x)$ is written in ____ form.

4. Create parametric equations $x = f(t)$, $y = g(t)$ and sketch their graph. Explain any interesting features of your graph based on the functions f and g.

Problems

In Exercises 5-8, sketch the graph of the given parametric equations by hand, making a table of points to plot. Be sure to indicate the orientation of the graph.

5. $x = t^2 + t, \quad y = 1 - t^2, \quad -3 \leq t \leq 3$

6. $x = 1, \quad y = 5\sin t, \quad -\pi/2 \leq t \leq \pi/2$

7. $x = t^2, \quad y = 2, \quad -2 \leq t \leq 2$

8. $x = t^3 - t + 3, \quad y = t^2 + 1, \quad -2 \leq t \leq 2$

In Exercises 9-17, sketch the graph of the given parametric equation; using a graphing utility is advisable. Be sure to indicate the orientation of the graph.

9. $x = t^3 - 2t^2, \quad y = t^2, \quad -2 \leq t \leq 3$

10. $x = 1/t, \quad y = \sin t, \quad 0 \leq t \leq 10$

11. $x = 3\cos t, \quad y = 5\sin t, \quad 0 \leq t \leq 2\pi$

12. $x = 3\cos t, \quad y = 5\sin t + 3, \quad 0 \leq t \leq 2\pi$

13. $x = \cos t, \quad y = \cos(2t), \quad 0 \leq t \leq \pi$

14. $x = \cos t, \quad y = \sin(2t), \quad 0 \leq t \leq 2\pi$

15. $x = 2\sec t, \quad y = 3\tan t, \quad -\pi/2 \leq t \leq \pi/2$

16. $x = \cos t + \frac{1}{4}\cos(8t), \quad y = \sin t + \frac{1}{4}\sin(8t), \quad 0 \leq t \leq 2\pi$

17. $x = \cos t + \frac{1}{4}\sin(8t), \quad y = \sin t + \frac{1}{4}\cos(8t), \quad 0 \leq t \leq 2\pi$

In Exercises 18-19, four sets of parametric equations are given. Describe how their graphs are similar and different. Be sure to discuss orientation and ranges.

18.
(a) $x = t \quad y = t^2, \quad -\infty < t < \infty$
(b) $x = \sin t \quad y = \sin^2 t, \quad -\infty < t < \infty$
(c) $x = e^t \quad y = e^{2t}, \quad -\infty < t < \infty$
(d) $x = -t \quad y = t^2, \quad -\infty < t < \infty$

19.
(a) $x = \cos t \quad y = \sin t, \quad 0 \leq t \leq 2\pi$
(b) $x = \cos(t^2) \quad y = \sin(t^2), \quad 0 \leq t \leq 2\pi$
(c) $x = \cos(1/t) \quad y = \sin(1/t), \quad 0 \leq t \leq 2\pi$
(d) $x = \cos(\cos t) \quad y = \sin(\cos t), \quad 0 \leq t \leq 2\pi$

In Exercises 20-29, eliminate the parameter in the given parametric equations.

20. $x = 2t + 5, \quad y = -3t + 1$

21. $x = \sec t, \quad y = \tan t$

22. $x = 4\sin t + 1, \quad y = 3\cos t - 2$

23. $x=t^2, \quad y=t^3$

24. $x=\frac{1}{t+1}, \quad y=\frac{3t+5}{t+1}$

25. $x=e^t, \quad y=e^{3t}-3$

26. $x=\ln t, \quad y=t^2-1$

27. $x=\cot t, \quad y=\csc t$

28. $x=\cosh t, \quad y=\sinh t$

29. $x=\cos(2t), \quad y=\sin t$

In Exercises 30-33, eliminate the parameter in the given parametric equations. Describe the curve defined by the parametric equations based on its rectangular form.

30. $x=at+x_0, \quad y=bt+y_0$

31. $x=r\cos t, \quad y=r\sin t$

32. $x=a\cos t+h, \quad y=b\sin t+k$

33. $x=a\sec t+h, \quad y=b\tan t+k$

In Exercises 34-37, find parametric equations for the given rectangular equation using the parameter $t=\frac{dy}{dx}$. Verify that at t $=1$, the point on the graph has a tangent line with slope of 1.

34. $y=3x^2-11x+2$

35. $y=e^x$

36. $y=\sin x$ on $[0,\pi]$

37. $y=\sqrt{x}$ on $[0,\pi]$

In Exercises 42-45, find the value(s) of *t* where the graph of the parametric equations crosses itself.

38. $x=t^3-t+3, \quad y=t^2-3$

39. $x=t^3-4t^2+t+7, \quad y=t^2-t$

40. $x=\cos t, \quad y=\sin(2t)$ on $[0,2\pi]$

41. $x=\cos t\cos(3t), \quad y=\sin t\cos(3t)$ on $[0,\pi]$

In Exercises 42-45, find the value(s) of *t* where the curve defined by the parametric equations is not smooth.

42. $x=t^3+t^2-t, \quad y=t^2+2t+3$

43. $x=t^2-4t, \quad y=t^3-2t^2-4t$

44. $x=\cos t, \quad y=2\cos t$

45. $x=2\cos t-\cos(2t), \quad y=2\sin t-\sin(2t)$

In Exercises 46-54, find parametric equations that describe the given situation.

46. A projectile is fired from a height of 0ft, landing 16ft away in 4s.

47. A projectile is fired from a height of 0ft, landing 200ft away in 4s.

48. A projectile is fired from a height of 0ft, landing 200ft away in 20s.

49. A circle of radius 2, centered at the origin, that is traced clockwise once on $[0,2\pi]$.

50. A circle of radius 3, centered at $(1,1)$, that is traced once counter-clockwise once on $[0,1]$.

51. An ellipse centered at (1,3) with vertical major axis of length 6 and minor axis of length 2.

52. An ellipse with foci at $(\pm 1,0)$ and vertices at $(\pm 5,0)$.

53. A hyperbola with foci at $(5,-3)$ and $(-1,-3)$, and with vertices at $(1,-3)$ and $(3,-3)$.

54. A hyperbola with vertices at $(0,\pm 6)$ and asymptotes $y=\pm 3x$.

9.3: Calculus and Parametric Equations

Terms and Concepts

1. T/F: Given parametric equations $x = f(t)$ and $y = g(t)$, $\frac{dy}{dx} = f'(t)f'(t)$, as long as $g'(t) \neq 0$.

2. Given parametric equations $x = f(t)$ and $y = g(t)$, the derivative $\frac{dy}{dx}$ as given in Key Idea 37 is a function of ________?

3. T/F: Given parametric equations $x = f(t)$ and $y = g(t)$, to find $\frac{d^2y}{dx^2}$, one simply computes $\frac{d}{dt}\left(\frac{dy}{dx}\right)$.

4. T/F: If $\frac{dy}{dx} = 0$ at $t = t_0$, then the normal line to the curve at $t = t_0$ is a vertical line.

Problems

In Exercises 5-12, parametric equations for a curve are given.
(a) Find $\frac{dy}{dx}$.
(b) Find the equations of the tangent and normal line(s) at the point(s) given.
(c) Sketch the graph of the parametric functions along with the found tangent and normal lines.

5. $x = t,\ y = t^2; \quad t = 1$

6. $x = \sqrt{t},\ y = 5t + 2; \quad t = 4$

7. $x = t^2 - t,\ y = t^2 + t; \quad t = 1$

8. $x = t^2,\ y = t^3 - 1; \quad t = 0$ and $t = 1$

9. $x = \sec t,\ y = \tan t$ on $(-\pi/2, \pi/2); \quad t = \pi/4$

10. $x = \cos t,\ y = \sin(2t); \quad t = \pi/4$

11. $x = \cos t \sin(2t),\ y = \sin t \sin(2t)$ on $[0, 2\pi]; \quad t = 3\pi/4$

12. $x = e^{t/10},\ y = e^{t/10} \sin t; \quad t = \pi/2$

In Exercises 13-20, find *t*-values where the curve defined by the given parametric equations has a horizontal tangent line. Note: these are the same equations as in Exercises 5-12.

13. $x = t,\ y = t^2$

14. $x = \sqrt{t},\ y = 5t + 2$

15. $x = t^2 - t,\ y = t^2 + t$

16. $x = t^2 - 1,\ y = t^3 - t$

17. $x = \sec t,\ y = \tan t$ on $(-\pi/2, \pi/2)$

18. $x = \cos t,\ y = \sin(2t)$ on $[0, 2\pi]$

19. $x = \cos t \sin(2t),\ y \sin t \sin(2t)$ on $[0, 2\pi]$

20. $x = e^{t/10} \cos t,\ y = e^{t/10} \sin t$

In Exercises 21-24, find $t = t_0$ where the graph of the given parametric equations is not smooth, then find $\lim\limits_{t \to t_0} \frac{dy}{dx}$.

21. $x = \frac{1}{t^2+1}, \quad y = t^3$

22. $x = -t^3 + 7t^2 - 16t + 13, \quad y = t^3 - 5t^2 + 8t - 2$

23. $x = t^3 - 3t^2 + 3t - 1, \quad y = t^2 - 2t + 1$

24. $x = \cos^2 t, \quad y = 1 - \sin^2 t$

In Exercises 25-32, parametric equations for a curve are given. Find $\frac{d^2y}{dx^2}$, then determine the intervals on which the graph of the curve is concave up/down. Note: these are the same equations as in Exercises 5-12.

25. $x = t,\ y = t^2; \quad t = 1$

26. $x = \sqrt{t},\ y = 5t + 2; \quad t = 4$

27. $x = t^2 - t,\ y = t^2 + t; \quad t = 1$

28. $x = t^2,\ y = t^3 - 1; \quad t = 0 \text{ and } t = 1$

29. $x = \sec t,\ y = \tan t \text{ on } (-\pi/2, \pi/2); \quad t = \pi/4$

30. $x = \cos t,\ y = \sin(2t); \quad t = \pi/4$

31. $x = \cos t \sin(2t),\ y = \sin t \sin(2t) \text{ on } [0, 2\pi]; \quad t = 3\pi/4$

32. $x = e^{t/10},\ y = e^{t/10} \sin t; \quad t = \pi/2$

In Exercises 37-40, numerically approximates the given arc length.

37. Approximate the arc length of one petal of the rose curve $x = \cos t \cos(2t)$ using Simpson's Rule and $n = 4$.

38. Approximate the arc length of the "bow tie curve" $x = \cos t, \quad y = \sin(2t)$ using Simpson's Rule and $n = 6$.

39. Approximate the arc length of the parabola $x = t^2 - t$, $y = t^2 + t$ on [-1,1] using Simpson's Rule and $n = 4$.

40. A common approximate of the circumference of an ellipse given by $x = a\cos t, \quad y = b\sin t$ is $C \approx 2\pi\sqrt{\frac{a^2+b^2}{2}}$.

In Exercises 41-44, a solid of revolution is described. Find or approximate its surface area as specified.

41. Find the surface area of the sphere formed by rotating the circle $x = 2\cos t, \quad y = 2\sin t$ about:
(a) the *x*-axis and
(b) the *y*-axis.

42. Find the surface are of the torus (or "donut") formed by rotating the circle $x = \cos t + 2, \quad y = \sin t$ about the *y*-axis.

43. Approximate the surface are of the solid formed by rotating the "upper right half" of the bow tie curve $x = \cos t, y = \sin(2t)$ on $[0, \pi/2]$ about the x-axis, using Simpson's Rule and $n = 4$.

44. Approximate the surface area of the solid formed by rotating the one petal of the rose curve $x = \cos t \cos(2t), \quad y = \sin t \cos(2t)$ on $[0, \pi/4]$ about the x-axis, using Simpson's Rule and $n = 4$.

9.4: Introduction to Polar Coordinates

Terms and Concepts

1. In your own words, describe how to plot the polar point $P(r, \theta)$.

2. T/F: When plotting a point with polar coordinate$P(r, \theta)$, r must be positive.

3. T/F: Every point in the Cartesian plane can be represented by a polar coordinate.

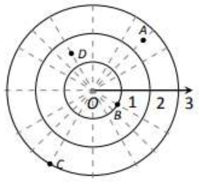
4. T/F: Every point in the Cartesian plane can be represented uniquely by a polar coordinate.

Problems

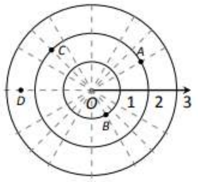
5. Plot the points with the given polar coordinates.
(a) $A = P(2, 0)$
(b) $B = P(1, \pi)$
(c) $C = P(-2, \pi/2)$
(d) $D = P(1, \pi/4)$

6. Plot the points with the given polar coordinates.
(a) $A = P(2, 3\pi)$
(b) $B = P(1, -\pi)$
(c) $C = P(1, 2)$
(d) $D = P(1/2, 5\pi/6)$

7. For each of the given points give two sets of polar coordinates that identify it, where $0 \le \theta \le 2\pi$.



8. For each of the given points give two sets of polar coordinates that identify it, where $-\pi \leq \theta \leq \pi$.



9. Convert each of the following polar coordinates to rectangular, and each of the following rectangular coordinates to polar.
(a) $A = P(2, \pi/4)$
(b) $B = P(2, -\pi/4)$
(c) $C = P(2, -1)$
(d) $D = P(-2, 1)$

10. Convert each of the following polar coordinates to rectangular, and each of the following rectangular coordinates to polar.
(a) $A = P(3, \pi)$
(b) $B = P(1, 2\pi/3)$
(c) $C = P(0, 4)$
(d) $D = P(1, -\sqrt{3})$

In Exercises 11-29, graph the polar function on the given interval.

11. $r = 2, \quad 0 \leq \theta \leq \pi/2$

12. $\theta = \pi/6, \quad -1 \leq r \leq \pi/2$

13. $r = 1 - \cos\theta, \quad [0, 2\pi]$

14. $r = 2 + \sin\theta, \quad [0, 2\pi]$

15. $r = 2 - \sin\theta, \quad [0, 2\pi]$

16. $r = 1 - 2\sin\theta, \quad [0, 2\pi]$

17. $r = 1 + 2\sin\theta, \quad [0, 2\pi]$

18. $r = \cos(2\theta), \quad [0, 2\pi]$

19. $r = \sin(3\theta), \quad [0, \pi]$

20. $r = \cos(\theta/3), \quad [0, 3\pi]$

21. $r = \cos(2\theta/3), \quad [0, 6\pi]$

22. $r = \theta/2, \quad [0, 4\pi]$

23. $r = 3\sin(\theta), \quad [0, \pi]$

24. $r = \cos\theta\sin\theta, \quad [0, 2\pi]$

25. $r = \theta^2 - (\pi/2)^2, \quad [\pi, \pi]$

26. $r = \frac{3}{5\sin\theta - \cos\theta}, \quad [0, 2\pi]$

27. $r = \frac{-2}{3\cos\theta - 2\sin\theta}, \quad [0, 2\pi]$

28. $r = 3\sec\theta, \quad (-\pi/2, \pi/2)$

29. $r = 3\csc\theta, \quad (0, \pi)$

In Exercises 30-38, convert the polar equation to a rectangular equation.

30. $r = 2\cos\theta$

31. $r = -4\sin\theta$

32. $r = \cos\theta + \sin\theta$

33. $r = \frac{7}{5\sin\theta - 2\cos\theta}$

34. $r = \frac{3}{\cos\theta}$

35. $r = \frac{4}{\sin\theta}$

36. $r = \tan\theta$

37. $r = 2$

38. $\theta = \pi/6$

In Exercises 39-46, convert the rectangular equation to a polar equation.

39. $y = x$

40. $y = 4x + 7$

41. $x = 5$

42. $y = 5$

43. $x = y^2$

44. $x^2y = 1$

45. $x^2 + y^2 = 7$

46. $x + 1)^2 + y^2 = 1$

In Exercises 47-54, find the points of intersection of the polar graphs.

47. $r = \sin(2\theta)$ and $r = \cos\theta$ on $[0, \pi]$.

48. $r = \cos(2\theta)$ and $r = \cos\theta$ on $[0, \pi]$.

49. $r = 2\cos\theta$ and $r = 2\sin\theta$ on $[0, \pi]$.

50. $r = \sin\theta$ and $r = \sqrt{3} + 3\sin\theta$ on $[0, 2\pi]$.

51. $r = \sin(3\theta)$ and $r = \cos(3\theta)$ on $[0, \pi]$.

52. $r = 3\cos\theta$ and $r = 1 + \cos\theta$ on $[-\pi, \pi]$.

53. $r = 1$ and $r = 2\sin(2\theta)$ on $[0, 2\pi]$.

54. $r = 1 - \cos\theta$ and $r = 1 + \sin\theta$ on $[0, 2\pi]$.

55. Pick an integer value for *n*, where $n \neq 2, 3$ and use technology to plot $r = \sin\left(\frac{m}{n}\theta\right)$ for three different integer values of *m*. Sketch these and determine a minimal interval on which the entire graph is shown.

56. Create your own polar function, $r = f(\theta)$ and sketch it. Describe why the graph looks as it does.

9.5: Calculus and Polar Functions

Terms and Concepts

1. Given polar equation $r = f(\theta)$, how can one create parametric equations of the same curve?

2. With rectangular coordinates, it is natural to approximate are with _________; with polar coordinates, it is natural to approximate area with ________.

Problems

In Exercises 3-10, find;
(a) $\frac{dy}{dx}$
(b) the equation of the tangent and normal lines to the curve at the indicated θ-value.

3. $r = 1; \quad \theta = \pi/4$

4. $r = \cos\theta; \quad \theta = \pi/4$

5. $r = 1 + \sin\theta; \quad \theta = \pi/6$

6. $r = 1 - \cos\theta; \quad \theta = 3\pi/4$

7. $r = \theta; \quad \theta = \pi/2$

8. $r = \cos(3\theta); \quad \theta = \pi/6$

9. $r = \sin(4\theta); \quad \theta = \pi/3$

10. $r = \frac{1}{\sin\theta - \cos\theta}; \quad \theta = \pi$

In Exercises 11-14, find the values of θ in the given interval where the graph of the polar function has horizontal and vertical tangent lines.

11. $r = 3; \quad [0, 2\pi]$

12. $r = 2\sin\theta; \quad [0, \pi]$

13. $r = \cos(2\theta); \quad [0, 2\pi]$

14. $r = 1 + \cos\theta; \quad [0, 2\pi]$

In Exercises 15-16, find the equation of the lines tangent to the graph at the pole.

15. $r = \sin\theta; \quad [0, \pi]$

16. $r = \sin(3\theta); \quad [0, \pi]$

In Exercises 17-27, find the area of the described region.

17. Enclosed by the circle: $r = 4\sin\theta$

18. Enclosed by the circle $r = 5$

19. Enclosed by one petal of $r = \sin(3\theta)$

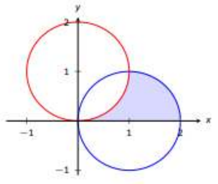
20. Enclosed by the cardiod $r = 1 - \sin\theta$

21. Enclosed by the inner loop of the limacon $r = 1 + 2\cos t$

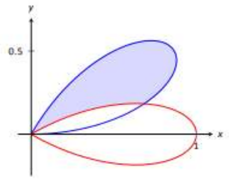
22. Enclosed by the outer loop of the limacon $r = 1 + 2\cos t$ (including area enclosed by the inner loop)

23. Enclosed between the inner and outer loop of the limacon $r = 1 + 2\cos t$

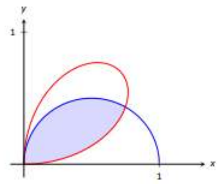
24. Enclosed by $r = 2\cos\theta$ and $r = \sin\theta$, as shown:



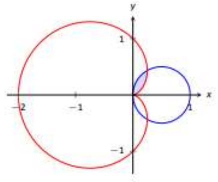
25. Enclosed by $r = \cos(3\theta)$ and $r = \sin(3\theta)$, as shown:



26. Enclosed by $r = \cos\theta$ and $r = \sin(2\theta)$, as shown:



27. Enclosed by $r=\cos\theta$ and $r=1-\cos\theta$, as shown:



In Exercises 28-32, answer the questions involving arc length.

28. Let $x(\theta)=f(\theta)\cos\theta$ and $y(\theta)=f(\theta)\sin\theta$. Show, as suggested by the text, that $x'(\theta)^2+y'(\theta)^2=f'(\theta)^2+f(\theta)^2$.

29. Use the arc length formula to compute the arc length of the circle $r=2$.

30. Use the arc length formula to compute the arc length of the circle $r=4\sin\theta$.

31. Approximate the arc length of one petal of the rose curve $r=\sin(3\theta)$ with Simpson's Rule and $n=4$.

32. Approximate the arc length of the cardiod $r=1+\cos\theta$ with Simpson's Rule and $n=6$.

In Exercises 33-37, answer the questions involving surface area.

33. Use Key Idea 44 to find the surface area of the sphere formed by revolving the circle $r=2$ about the initial ray.

34. Use Key Idea 44 to find the surface area of the sphere formed by revolving the circle $r=2\cos\theta$ about the initial ray.

35. Find the surface area of a solid formed by revolving the cardiod $r=1+\cos\theta$ about the initial ray.

36. Find the surface area of the solid formed by revolving the circle $r=2\cos\theta$ about the line $\theta=\pi/2$.

37. Find the surface area of the solid formed by revolving the line $r=3\sec\theta$, $-\pi/4\leq\theta\leq\pi/4$, about the line $\theta=\pi/2$.

This page titled 9.E: Applications of Curves in a Plane (Exercises) is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

10: Vectors

This chapter introduces a new mathematical object, the **vector**. We will see that vectors provide a powerful language for describing quantities that have magnitude and direction aspects. A simple example of such a quantity is force: when applying a force, one is generally interested in how much force is applied (i.e., the magnitude of the force) and the direction in which the force was applied. Vectors will play an important role in many of the subsequent chapters in this text. This chapter begins with moving our mathematics out of the plane and into "space." That is, we begin to think mathematically not only in two dimensions, but in three. With this foundation, we can explore vectors both in the plane and in space.

10.1: Introduction to Cartesian Coordinates in Space
10.2: An Introduction to Vectors
10.3: The Dot Product
10.4: The Cross Product
10.5: Lines
10.6: Planes
10.E: Applications of Vectors (Exercises)

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 10: Vectors is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

10.1: Introduction to Cartesian Coordinates in Space

Up to this point in this text we have considered mathematics in a 2--dimensional world. We have plotted graphs on the x-y plane using rectangular and polar coordinates and found the area of regions in the plane. We have considered properties of *solid* objects, such as volume and surface area, but only by first defining a curve in the plane and then rotating it out of the plane.

While there is wonderful mathematics to explore in "2D," we live in a "3D" world and eventually we will want to apply mathematics involving this third dimension. In this section we introduce Cartesian coordinates in space and explore basic surfaces. This will lay a foundation for much of what we do in the remainder of the text.

Each point P in space can be represented with an ordered triple, $P=(a,b,c)$, where a, b and c represent the relative position of P along the x-, y- and z-axes, respectively. Each axis is perpendicular to the other two.

Visualizing points in space on paper can be problematic, as we are trying to represent a 3-dimensional concept on a 2--dimensional medium. We cannot draw three lines representing the three axes in which each line is perpendicular to the other two. Despite this issue, standard conventions exist for plotting shapes in space that we will discuss that are more than adequate.

One convention is that the axes must conform to the **right hand rule**. This rule states that when the index finger of the right hand is extended in the direction of the positive x-axis, and the middle finger (bent "inward" so it is perpendicular to the palm) points along the positive y-axis, then the extended thumb will point in the direction of the positive z-axis. (It may take some thought to verify this, but this system is inherently different from the one created by using the "left hand rule."). There are two popular methods to draw axes that we briefly discuss.

In Figure $10.1.1$ we see the point $P=(2,1,3)$ plotted on a set of axes. The basic convention here is that the x-y plane is drawn in its standard way, with the z-axis down to the left. The perspective is that the paper represents the x-y plane and the positive z axis is coming up, off the page. This method is preferred by many engineers. Because it can be hard to tell where a single point lies in relation to all the axes, dashed lines have been added to let one see how far along each axis the point lies.

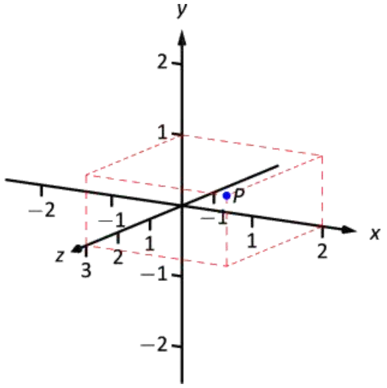


Figure $10.1.1$: Plotting the point $P=(2,1,3)$ in space.

One can also consider the x-y plane as being a horizontal plane in, say, a room, where the positive z-axis is pointing up. When one steps back and looks at this room, one might draw the axes as shown in Figure $10.1.2$. The same point P is drawn, again with dashed lines. This point of view is preferred by most mathematicians, and is the convention adopted by this text.

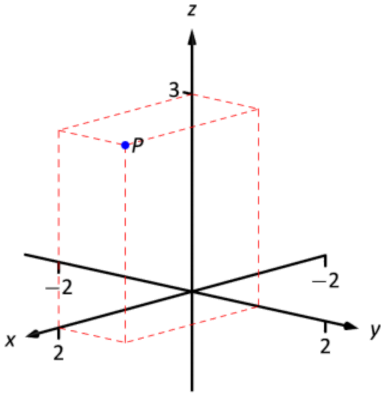


Figure $10.1.2$: Plotting the point $P=(2,1,3)$ in space with a perspective used in this text.

Note

As long as the coordinate axes are positioned so that they follow the right hand rule, it does not matter how the axes are drawn on paper.

Measuring Distances

It is of critical importance to know how to measure distances between points in space. The formula for doing so is based on measuring distance in the plane, and is known (in both contexts) as the Euclidean measure of distance.

Definition 48: distance in space

Let $P=(x_1,y_1,z_1)$ and $Q=(x_2,y_2,z_2)$ be points in space. The distance D between P and Q is

$$D=\sqrt{(x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2}. \tag{10.1.1}$$

We refer to the line segment that connects points P and Q in space as $\overline{PQ}$, and refer to the length of this segment as $||\overline{PQ}||$. The above distance formula allows us to compute the length of this segment.

Example 10.1.1: Length of a line segment

Let $P=(1,4,-1)$ and let $Q=(2,1,1)$. Draw the line segment $\overline{PQ}$ and find its length.

Solution

The points P and Q are plotted in Figure $10.1.3$; no special consideration need be made to draw the line segment connecting these two points; simply connect them with a straight line.

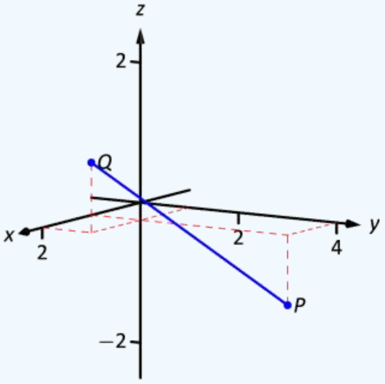


Figure $10.1.3$: Plotting points P and Q in Example $10.1.1$.

One *cannot* actually measure this line on the page and deduce anything meaningful; its true length must be measured analytically. Applying Definition 48, we have

$$||\overline{PQ}||=\sqrt{(2-1)^2+(1-4)^2+(1-(-1))^2}=\sqrt{14}\approx 3.74. \tag{10.1.2}$$

Spheres

Just as a circle is the set of all points in the *plane* equidistant from a given point (its center), a sphere is the set of all points in *space* that are equidistant from a given point. Definition 48 allows us to write an equation of the sphere. We start with a point $C=(a,b,c)$ which is to be the center of a sphere with radius r. If a point $P=(x,y,z)$ lies on the sphere, then P is r units from C; that is,

$$||\overline{PC}||=\sqrt{(x-a)^2+(y-b)^2+(z-c)^2}=r. \tag{10.1.3}$$

Squaring both sides, we get the standard equation of a sphere in space with center at $C=(a,b,c)$ with radius r, as given in the following Key Idea.

KEY IDEA 45: STANDARD EQUATION OF A SPHERE IN SPACE

The standard equation of the sphere with radius r, centered at $C = (a, b, c)$, is

$$(x-a)^2 + (y-b)^2 + (z-c)^2 = r^2. \tag{10.1.4}$$

Example 10.1.2: Equation of a sphere

Find the center and radius of the sphere defined by $x^2 + 2x + y^2 - 4y + z^2 - 6z = 2$.

Solution

To determine the center and radius, we must put the equation in standard form. This requires us to complete the square (three times).

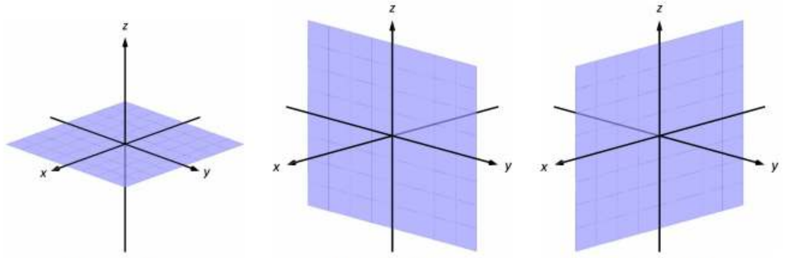
$$\begin{aligned} x^2 + 2x + y^2 - 4y + z^2 - 6z &= 2 \\ (x^2 + 2x + 1) + (y^2 - 4y + 4) + (z^2 - 6z + 9) - 14 &= 2 \\ (x+1)^2 + (y-2)^2 + (z-3)^2 &= 16 \end{aligned}$$

The sphere is centered at $(-1, 2, 3)$ and has a radius of 4.

The equation of a sphere is an example of an implicit function defining a surface in space. In the case of a sphere, the variables x, y and z are all used. We now consider situations where surfaces are defined where one or two of these variables are absent.

Introduction to Planes in Space

The coordinate axes naturally define three planes (shown in Figure $10.1.4$), the **coordinate planes**: the x-y plane, the y-z plane and the x-z plane. The x-y plane is characterized as the set of all points in space where the z-value is 0. This, in fact, gives us an equation that describes this plane: $z = 0$. Likewise, the x-z plane is all points where the y-value is 0, characterized by $y = 0$.



the *x*-*y* plane the *y*-*z* plane the *x*-*z* plane

Figure $10.1.4$: The coordinate planes.

The equation $x = 2$ describes all points in space where the x-value is 2. This is a plane, parallel to the y-z coordinate plane, shown in Figure $10.1.5$.

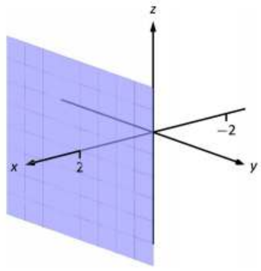


Figure $10.1.5$: The plane $x = 2$.

Example $10.1.3$: Regions defined by planes

Sketch the region defined by the inequalities $-1 \leq y \leq 2$.

Solution

The region is all points between the planes $y=-1$ and $y=2$. These planes are sketched in Figure $10.1.6$, which are parallel to the x-z plane. Thus the region extends infinitely in the x and z directions, and is bounded by planes in the y direction.

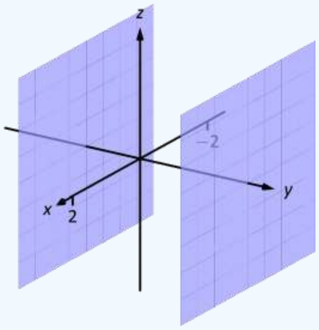


Figure $10.1.6$: Sketching the boundaries of a region in Example $10.1.3$.

Cylinders

The equation $x=1$ obviously lacks the y and z variables, meaning it defines points where the y and z coordinates can take on any value. Now consider the equation $x^2+y^2=1$ *in space*. In *the plane*, this equation describes a circle of radius 1, centered at the origin. In space, the z coordinate is not specified, meaning it can take on any value. In Figure $10.1.7a$, we show part of the graph of the equation $x^2+y^2=1$ by sketching 3 circles: the bottom one has a constant z-value of -1.5, the middle one has a z-value of 0 and the top circle has a z-value of 1. By plotting *all* possible z-values, we get the surface shown in Figure $10.1.7b$.
This surface looks like a "tube," or a "cylinder"; mathematicians call this surface a **cylinder** for an entirely different reason.

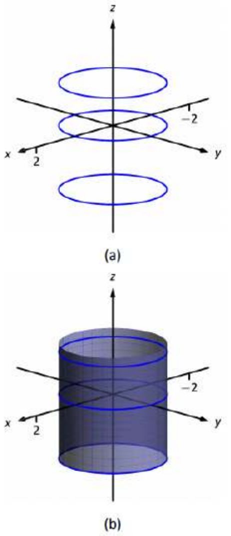


Figure $10.1.7$: Sketching $x^2+y^2=1$.

Definition 49: CYLINDER, Directrix and Rulings

Let C be a curve in a plane and let L be a line not parallel to C. A **cylinder** is the set of all lines parallel to L that pass through C. The curve C is the **directrix** of the cylinder, and the lines are the **rulings**.

In this text, we consider curves C that lie in planes parallel to one of the coordinate planes, and lines L that are perpendicular to these planes, forming **right cylinders**. Thus the directrix can be defined using equations involving 2 variables, and the rulings will be parallel to the axis of the 3rd variable.

In the example preceding the definition, the curve $x^2+y^2=1$ in the x-y plane is the directrix and the rulings are lines parallel to the z-axis. (Any circle shown in Figure 10.8 can be considered a directrix; we simply choose the one where $z=0$.) Sample rulings can also be viewed in part (b) of the figure. More examples will help us understand this definition.

Example $10.1.4$: Graphing cylinders

Graph the cylinder following cylinders.

1. $z=y^2$
2. $x=\sin z$

Solution

1. We can view the equation $z=y^2$ as a parabola in the y-z plane, as illustrated in Figure $10.1.8a$. As x does not appear in the equation, the rulings are lines through this parabola parallel to the x-axis, shown in Figure $10.1.8b$. These rulings give a general idea as to what the surface looks like, drawn in (c).

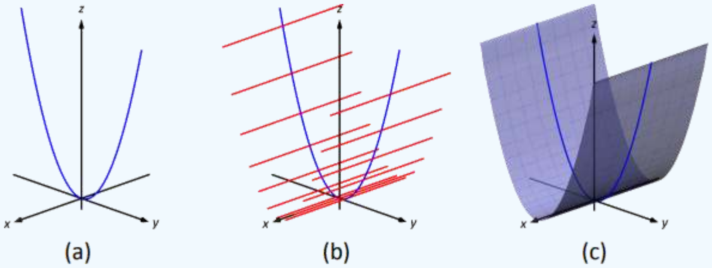


Figure $10.1.8$: Sketching the cylinder defined by $z=y^2$.

2. We can view the equation $x=\sin z$ as a sine curve that exists in the x-z plane, as shown in Figure $10.1.9a$. The rules are parallel to the y axis as the variable y does not appear in the equation $x=\sin z$; some of these are shown in Figure $10.1.9b$. The surface is shown in part (c) of the figure.

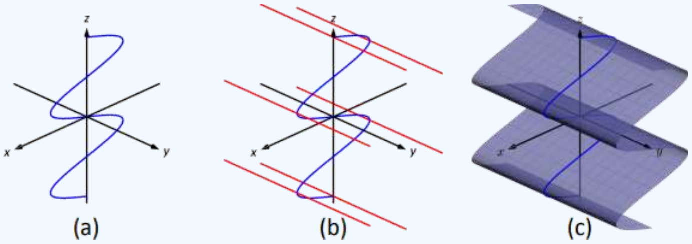


Figure $10.1.9$: Sketching the cylinder defined by $x=\sin z$.

Surfaces of Revolution

One of the applications of integration we learned previously was to find the volume of solids of revolution -- solids formed by revolving a curve about a horizontal or vertical axis. We now consider how to find the equation of the surface of such a solid.

Consider the surface formed by revolving $y=\sqrt{x}$ about the x-axis. Cross--sections of this surface parallel to the y-z plane are circles, as shown in Figure $10.1.1a$. Each circle has equation of the form $y^2+z^2=r^2$ for some radius r. The radius is a function of x; in fact, it is $r(x)=\sqrt{x}$. Thus the equation of the surface shown in Figure $10.1.10b$ is $y^2+z^2=(\sqrt{x})^2$.

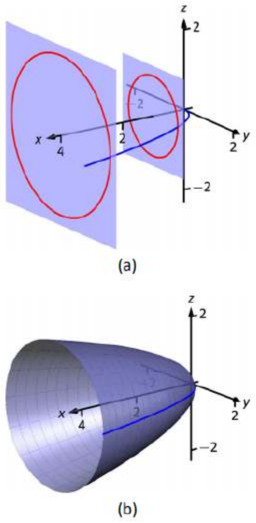


Figure 10.1.10: Introducing surfaces of revolution.

We generalize the above principles to give the equations of surfaces formed by revolving curves about the coordinate axes.

KEY IDEA 46: SURFACES OF REVOLUTION, PART 1

Let r be a radius function.

1. The equation of the surface formed by revolving $y=r(x)$ or $z=r(x)$ about the x-axis is $y^2+z^2=r(x)^2$.
2. The equation of the surface formed by revolving $x=r(y)$ or $z=r(y)$ about the y-axis is $x^2+z^2=r(y)^2$.
3. The equation of the surface formed by revolving $x=r(z)$ or $y=r(z)$ about the z-axis is $x^2+y^2=r(z)^2$.

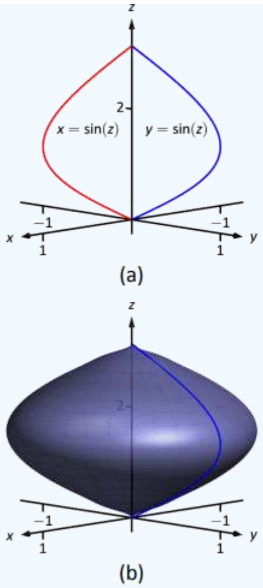
Example 10.1.5: Finding equation of a surface of revolution

Let $y=\sin z$ on $[0,\pi]$. Find the equation of the surface of revolution formed by revolving $y=\sin z$ about the z-axis.

Solution

Using Key Idea 46, we find the surface has equation $x^2+y^2=\sin^2 z$. The curve is sketched in Figure 10.1.11a and the surface is drawn in Figure 10.1.11b.

Note how the surface (and hence the resulting equation) is the same if we began with the curve $x=\sin z$, which is also drawn in Figure 10.1.11a.



(a)

(b)

Figure $10.1.11$: Revolving $y=\sin z$ about the z-axis in Example $10.1.5$.

This particular method of creating surfaces of revolution is limited. For instance, in Section 7.3 we found the volume of the solid formed by revolving $y=\sin x$ about the y-axis. Our current method of forming surfaces can only rotate $y=\sin x$ about the x-axis. Trying to rewrite $y=\sin x$ as a function of y is not trivial, as simply writing $x=\sin^{-1} y$ only gives part of the region we desire.

What we desire is a way of writing the surface of revolution formed by rotating $y=f(x)$ about the y-axis. We start by first recognizing this surface is the same as revolving $z=f(x)$ about the z-axis. This will give us a more natural way of viewing the surface.

A value of x is a measurement of distance from the z-axis. At the distance r, we plot a z-height of $f(r)$. When rotating $f(x)$ about the z-axis, we want all points a distance of r from the z-axis in the x-y plane to have a z-height of $f(r)$. All such points satisfy the equation $r^2=x^2+y^2$; hence $r=\sqrt{x^2+y^2}$. Replacing r with $\sqrt{x^2+y^2}$ in $f(r)$ gives $z=f(\sqrt{x^2+y^2})$. This is the equation of the surface.

KEY IDEA 47: SURFACES OF REVOLUTION, PART 2

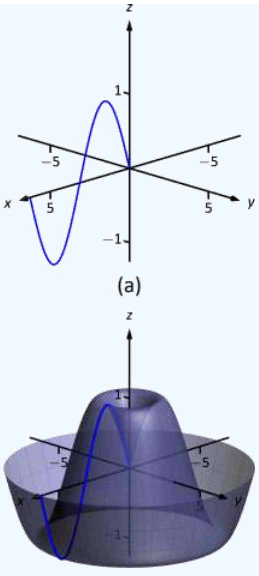
Let $z=f(x)$, $x\geq 0$, be a curve in the x-z plane. The surface formed by revolving this curve about the z-axis has equation $z=f\left(\sqrt{x^2+y^2}\right)$.

Example $10.1.6$: Finding equation of surface of revolution

Find the equation of the surface found by revolving $z=\sin x$ about the z-axis.

Solution

Using Key Idea 47, the surface has equation $z=\sin\left(\sqrt{x^2+y^2}\right)$. The curve and surface are graphed in Figure $10.1.12$.



(a)

(b)

Figure 10.1.12: Revolving $z=\sin x$ about the z-axis in Example 10.1.6.

Quadric Surfaces

Spheres, planes and cylinders are important surfaces to understand. We now consider one last type of surface, a **quadric surface**. The definition may look intimidating, but we will show how to analyze these surfaces in an illuminating way.

Definition 50: QUADRIC SURFACE

A **quadric surface** is the graph of the general second--degree equation in three variables:

$$Ax^2+By^2+Cz^2+Dxy+Exz+Fyz+Gx+Hy+Iz+J=0. \tag{10.1.5}$$

When the coefficients D, E or F are not zero, the basic shapes of the quadric surfaces are rotated in space. We will focus on quadric surfaces where these coeffiecients are 0; we will not consider rotations. There are six basic quadric surfaces: the elliptic paraboloid, elliptic cone, ellipsoid, hyperboloid of one sheet, hyperboloid of two sheets, and the hyperbolic paraboloid.

We study each shape by considering **traces**, that is, intersections of each surface with a plane parallel to a coordinate plane. For instance, consider the elliptic paraboloid $z=x^2/4+y^2$, shown in Figure 10.13. If we intersect this shape with the plane $z=d$ (i.e., replace z with d), we have the equation:

$$d=\frac{x^2}{4}+y^2. \tag{10.1.6}$$

Divide both sides by d:

$$1=\frac{x^2}{4d}+\frac{y^2}{d}. \tag{10.1.7}$$

This describes an ellipse -- so cross sections parallel to the x-y coordinate plane are ellipses. This ellipse is drawn in Figure 10.1.13.

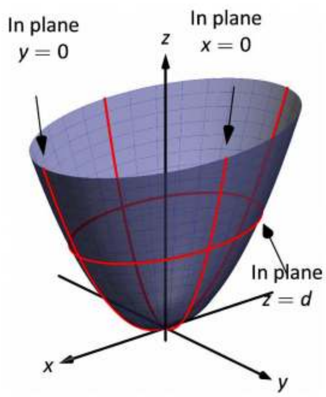


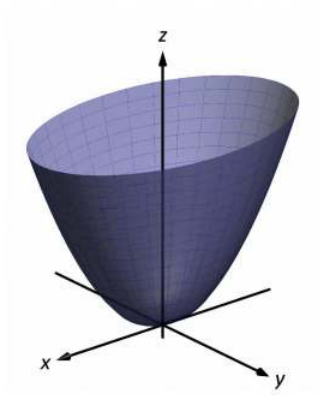
Figure 10.1.13: The elliptic paraboloid $z = x^2/4 + y^2$.

Now consider cross sections parallel to the x-z plane. For instance, letting $y = 0$ gives the equation $z = x^2/4$, clearly a parabola. Intersecting with the plane $x = 0$ gives a cross section defined by $z = y^2$, another parabola. These parabolas are also sketched in the figure.

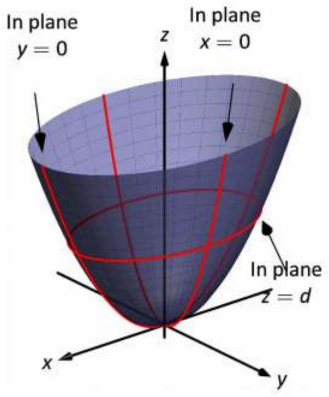
Thus we see where the elliptic paraboloid gets its name: some cross sections are ellipses, and others are parabolas.

Such an analysis can be made with each of the quadric surfaces. We give a sample equation of each, provide a sketch with representative traces, and describe these traces.

Elliptic Paraboloid, $z = \frac{x^2}{a^2} + \frac{y^2}{b^2}$



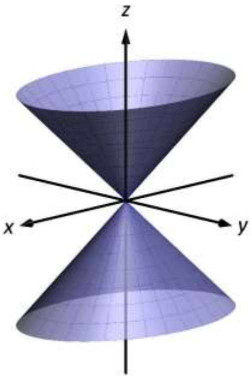
Plane	Trace
$x = d$	Parabola
$y = d$	Parabola
$z = d$	Ellipse



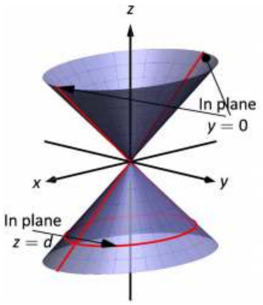
One variable in the equation of the elliptic paraboloid will be raised to the first power; above, this is the z variable. The paraboloid will "open" in the direction of this variable's axis. Thus $x = y^2/a^2 + z^2/b^2$ is an elliptic paraboloid that opens along the x-axis.

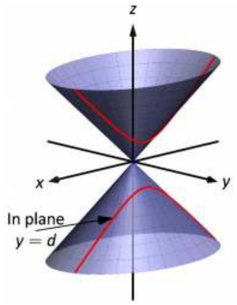
Multiplying the right hand side by (-1) defines an elliptic paraboloid that "opens" in the opposite direction.

Elliptic Cone, $z^2 = \frac{x^2}{a^2} + \frac{y^2}{b^2}$



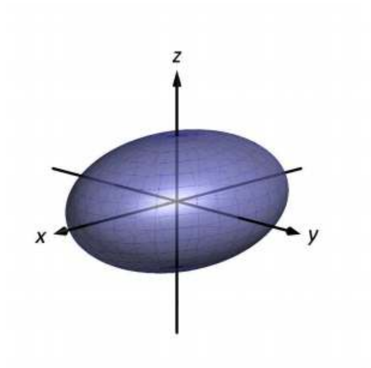
Plane	Trace
$x = 0$	Crossed Lines
$y = 0$	Crossed Lines
$x = d$	Hyperbola
$y = d$	Hyperbola
$z = d$	Ellipse



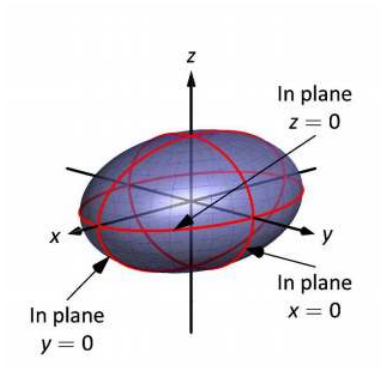


One can rewrite the equation as $z^2 - x^2/a^2 - y^2/b^2 = 0$. The one variable with a positive coefficient corresponds to the axis that the cones "open" along.

Ellipsoid, $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$

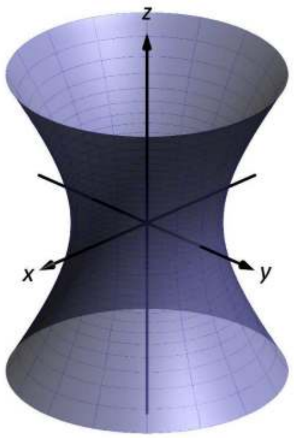


Plane	Trace
$x = d$	Ellipse
$y = d$	Ellipse
$z = d$	Ellipse

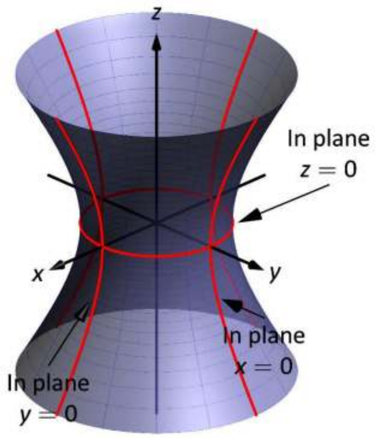

z
In plane
$z = 0$
x
y
In plane
$x = 0$
In plane
$y = 0$

If $a = b = c \neq 0$, the ellipsoid is a sphere with radius a; compare to Key Idea 45.

Hyperboloid of One Sheet, $\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$

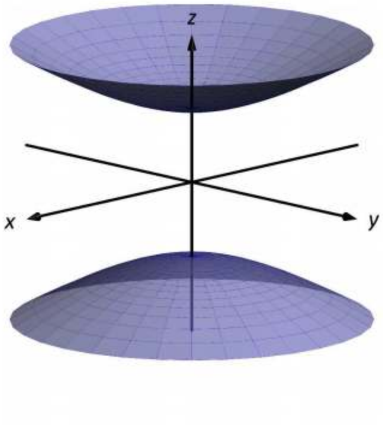


Plane	Trace
$x = d$	Hyperbola
$y = d$	Hyperbola
$z = d$	Ellipse

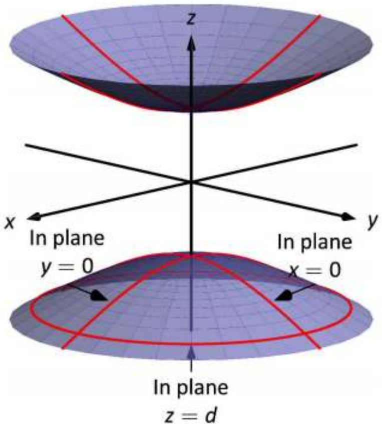

z
In plane
$z = 0$
x
y
In plane
$x = 0$
In plane
$y = 0$

The one variable with a negative coefficient corresponds to the axis that the hyperboloid "opens" along.

Hyperboloid of Two Sheets, $\frac{z^2}{c^2} - \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$

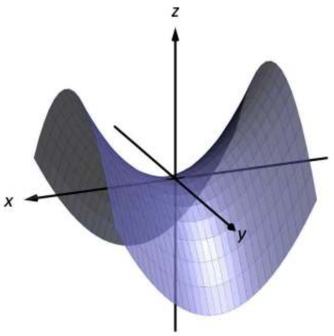


Plane	Trace
$x = d$	Hyperbola
$y = d$	Hyperbola
$z = d$	Ellipse

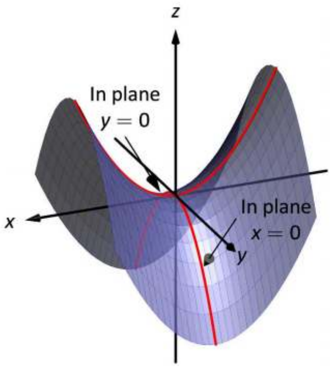

z
x
y
In plane
$y = 0$
In plane
$x = 0$
In plane
$z = d$

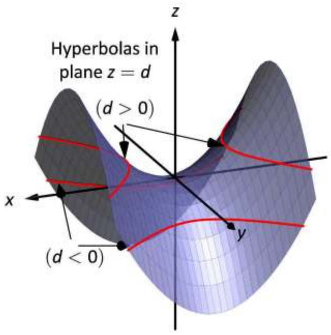
The one variable with a positive coefficient corresponds to the axis that the hyperboloid "opens" along. In the case illustrated, when $|d| < |c|$, there is no trace.

Hyperbolic Paraboloid, $z=\frac{x^2}{a^2}-\frac{y^2}{b^2}$



Plane	Trace
$x = d$	Parabola
$y = d$	Parabola
$z = d$	Hyperbola





The parabolic traces will open along the axis of the one variable that is raised to the first power.

Example 10.1.7: Sketching quadric surfaces

Sketch the quadric surface defined by the given equation.

1. $y=\frac{x^2}{4}+\frac{z^2}{16}$
2. $x^2+\frac{y^2}{9}+\frac{z^2}{4}=1.$
3. $z=y^2-x^2.$

Solution

1. $y=\frac{x^2}{4}+\frac{z^2}{16}$:

   We first identify the quadric by pattern--matching with the equations given previously. Only two surfaces have equations where one variable is raised to the first power, the elliptic paraboloid and the hyperbolic paraboloid. In the latter case, the other variables have different signs, so we conclude that this describes a hyperbolic paraboloid. As the variable with the first power is y, we note the paraboloid opens along the y-axis.

   To make a decent sketch by hand, we need only draw a few traces. In this case, the traces $x=0$ and $z=0$ form parabolas that outline the shape.

   $x=0$: The trace is the parabola $y=z^2/16$

   $z=0$: The trace is the parabola $y=x^2/4$.

   Graphing each trace in the respective plane creates a sketch as shown in Figure $10.1.14a$. This is enough to give an idea of what the paraboloid looks like. The surface is filled in in Figure $10.1.14b$.

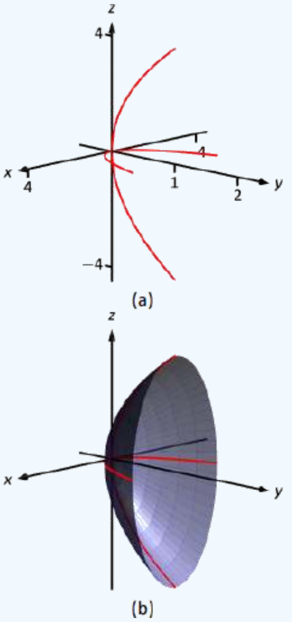


Figure 10.1.14***:*** *Sketching an elliptic paraboloid.*

2. $x^2+\frac{y^2}{9}+\frac{z^2}{4}=1$:

This is an ellipsoid. We can get a good idea of its shape by drawing the traces in the coordinate planes.

$x=0$: The trace is the ellipse $\frac{y^2}{9}+\frac{z^2}{4}=1$. The major axis is along the y--axis with length 6 (as $b=3$, the length of the axis is 6); the minor axis is along the z-axis with length 4.

$y=0$: The trace is the ellipse $x^2+\frac{z^2}{4}=1$. The major axis is along the z-axis, and the minor axis has length 2 along the x-axis.

$z=0$: The trace is the ellipse $x^2+\frac{y^2}{9}=1$, with major axis along the y-axis.

Graphing each trace in the respective plane creates a sketch as shown in Figure Figure $10.1.15a$. Filling in the surface gives Figure Figure $10.1.15b$.

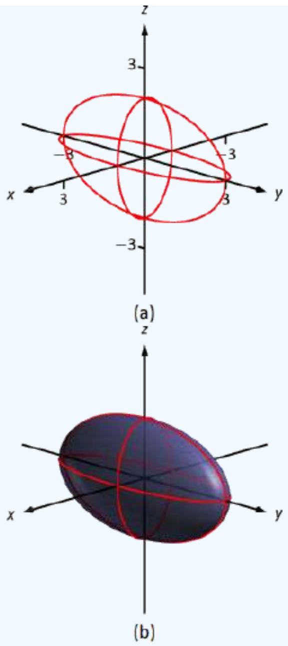


Figure 10.1.15***:*** *Sketching an ellipsoid.*

3. $z = y^2 - x^2$:

This defines a hyperbolic paraboloid, very similar to the one shown in the gallery of quadric sections. Consider the traces in the $y-z$ and $x-z$ planes:

$x = 0$: The trace is $z = y^2$, a parabola opening up in the $y-z$ plane.

$y = 0$: The trace is $z = -x^2$, a parabola opening down in the $x-z$ plane.

Sketching these two parabolas gives a sketch like that in Figure Figure $10.1.16a$, and filling in the surface gives a sketch like Figure $10.1.16b$.

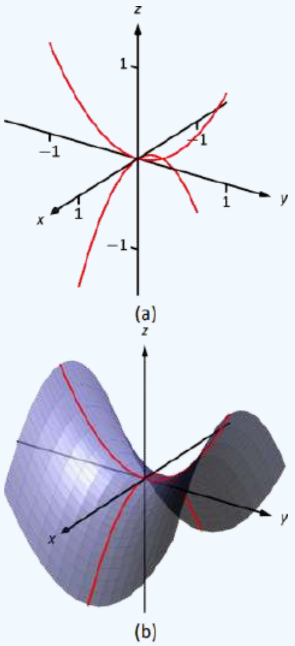


Figure 10.1.16: Sketching a hyperbolic paraboloid.

Example 10.1.8: Identifying quadric surfaces

Consider the quadric surface shown in Figure Figure 10.1.17. Which of the following equations best fits this surface?

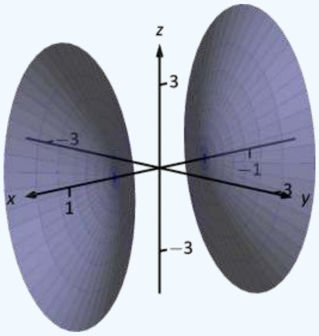


Figure 10.1.8.

$$(a)\, x^2-y^2-\frac{z^2}{9}=0 \qquad (c)\, z^2-x^2-y^2=1 \qquad (10.1.8)$$
$$(b)\, x^2-y^2-z^2=1 \qquad (d)\, 4x^2-y^2-\frac{z^2}{9}=1$$

Solution

The image clearly displays a hyperboloid of two sheets. The gallery informs us that the equation will have a form similar to $\frac{z^2}{c^2}-\frac{x^2}{a^2}-\frac{y^2}{b^2}=1$.

We can immediately eliminate option (a), as the constant in that equation is not 1.

The hyperboloid "opens" along the x-axis, meaning x must be the only variable with a positive coefficient, eliminating (c).

The hyperboloid is wider in the z-direction than in the y-direction, so we need an equation where $c>b$. This eliminates (b), leaving us with (d). We should verify that the equation given in (d), $4x^2-y^2-\frac{z^2}{9}=1$, fits.

We already established that this equation describes a hyperboloid of two sheets that opens in the x-direction and is wider in the z-direction than in the y. Now note the coefficient of the x-term. Rewriting $4x^2$ in standard form, we have: $4x^2=\frac{x^2}{(1/2)^2}$. Thus when $y=0$ and $z=0$, x must be $1/2$; i.e., each hyperboloid "starts" at $x=1/2$. This matches our figure.

We conclude that $4x^2-y^2-\frac{z^2}{9}=1$ best fits the graph.

This section has introduced points in space and shown how equations can describe surfaces. The next sections explore *vectors*, an important mathematical object that we'll use to explore curves in space.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 10.1: Introduction to Cartesian Coordinates in Space is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

10.2: An Introduction to Vectors

Many quantities we think about daily can be described by a single number: temperature, speed, cost, weight and height. There are also many other concepts we encounter daily that cannot be described with just one number. For instance, a weather forecaster often describes wind with its speed and its direction ("... with winds from the southeast gusting up to 30 mph ..."). When applying a force, we are concerned with both the magnitude and direction of that force. In both of these examples, *direction* is important. Because of this, we study *vectors*, mathematical objects that convey both magnitude and direction information.

One "bare--bones" definition of a vector is based on what we wrote above: "a vector is a mathematical object with magnitude and direction parameters." This definition leaves much to be desired, as it gives no indication as to how such an object is to be used. Several other definitions exist; we choose here a definition rooted in a geometric visualization of vectors. It is very simplistic but readily permits further investigation.

Definition 51 Vector

A **vector** is a directed line segment.

Given points P and Q (either in the plane or in space), we denote with $\vec{PQ}$ the vector *from* P to Q. The point P is said to be the initial point of the vector, and the point Q is the **terminal point**.

The **magnitude**, **length** or **norm** of $\vec{PQ}$ is the length of the line segment $\overline{PQ}$:

$$\|\vec{PQ}\| = \|\overline{PQ}\|. \tag{10.2.1}$$

Two vectors are **equal** if they have the same magnitude and direction.

Figure 10.18 shows multiple instances of the same vector. Each directed line segment has the same direction and length (magnitude), hence each is the same vector.

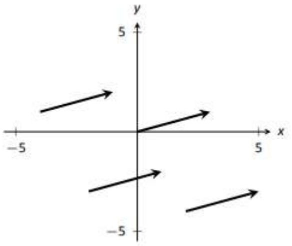


Figure 10.18: Drawing the same vector with different initial points.

We use $\mathbb{R}^2$ (pronounced "r two") to represent all the vectors in the plane, and use $\mathbb{R}^3$ (pronounced "r three") to represent all the vectors in space.

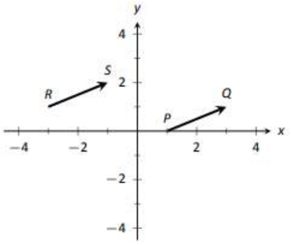


Figure 10.19: Illustrating how equal vectors have the same displacement.

Consider the vectors $\vec{PQ}$ and $\vec{RS}$ as shown in Figure 10.19. The vectors look to be equal; that is, they seem to have the same length and direction. Indeed, they are. Both vectors move 2 units to the right and 1 unit up from the initial point to reach the terminal point. One can analyze this movement to measure the magnitude of the vector, and the movement itself gives direction information (one could also measure the slope of the line passing through P and Q or R and S). Since they have the same length and direction, these two vectors are equal.

This demonstrates that inherently all we care about is *displacement*; that is, how far in the x, y and possibly z directions the terminal point is from the initial point. Both the vectors $\vec{PQ}$ and $\vec{RS}$ in Figure 10.19 have an x-displacement of 2 and a y-displacement of 1. This suggests a standard way of describing vectors in the plane. A vector whose x-displacement is a and whose y-displacement is b will have terminal point (a, b) when the initial point is the origin, $(0, 0)$. This leads us to a definition of a standard and concise way of referring to vectors.

Definition 52 Component Form of a Vector

1. The **component form** of a vector $\vec{v}$ in $\mathbb{R}^2$, whose terminal point is $(a,\ b)$ when its initial point is $(0,\ 0)$, is $\langle a, b\rangle$.
2. The **component form** of a vector $\vec{v}$ in $\mathbb{R}^3$, whose terminal point is $(a,\ b,\ c)$ when its initial point is $(0,\ 0,\ 0)$, is $\langle a, b, c\rangle$.

The numbers a, b (and c, respectively) are the **components** of $\vec{v}$.

It follows from the definition that the component form of the vector $\vec{PQ}$, where $P = (x_1, y_1)$ and $Q = (x_2, y_2)$ is

$$\vec{PQ} = \langle x_2 - x_1, y_2 - y_1\rangle; \tag{10.2.2}$$

in space, where $P = (x_1, y_1, z_1)$ and $Q = (x_2, y_2, z_2)$, the component form of $\vec{PQ}$ is

$$\vec{PQ} = \langle x_2 - x_1, y_2 - y_1, z_2 - z_1\rangle. \tag{10.2.3}$$

We practice using this notation in the following example.

Example 10.2.1: Using component form notation for vectors

1. Sketch the vector $\vec{v} = \langle 2, -1\rangle$ starting at $P = (3, 2)$ and find its magnitude.
2. Find the component form of the vector $\vec{w}$ whose initial point is $R = (-3, -2)$ and whose terminal point is $S = (-1, 2)$.
3. Sketch the vector $\vec{u} = \langle 2, -1, 3\rangle$ starting at the point $Q = (1, 1, 1)$ and find its magnitude.

Solution

1. Using P as the initial point, we move 2 units in the positive x-direction and -1 units in the positive y-direction to arrive at the terminal point $P\,' = (5, 1)$, as drawn in Figure 10.20(a).
   The magnitude of $\vec{v}$ is determined directly from the component form:
   $$\|\vec{v}\| = \sqrt{2^2 + (-1)^2} = \sqrt{5}. \tag{10.2.4}$$
2. Using the note following Definition 52, we have
   $$\vec{RS} = \langle -1 - (-3), 2 - (-2)\rangle = \langle 2, 4\rangle. \tag{10.2.5}$$

   One can readily see from Figure 10.20(a) that the x- and y-displacement of $\vec{RS}$ is 2 and 4, respectively, as the component form suggests.
3. Using Q as the initial point, we move 2 units in the positive x-direction, -1 unit in the positive y-direction, and 3 units in the positive z-direction to arrive at the terminal point $Q' = (3, 0, 4)$, illustrated in Figure 10.20(b).

   The magnitude of $\vec{u}$ is:
   $$\|\vec{u}\| = \sqrt{2^2 + (-1)^2 + 3^2} = \sqrt{14}. \tag{10.2.6}$$

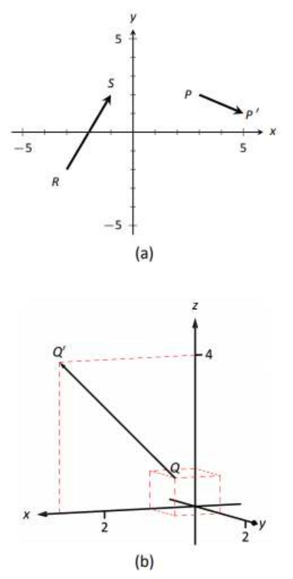


Figure 10.20: Graphing vectors in Example 10.2.1

Now that we have defined vectors, and have created a nice notation by which to describe them, we start considering how vectors interact with each other. That is, we define an *algebra* on vectors.

Definition 53 VECTOR ALGEBRA

1. Let $\vec{u} = \langle u_1, u_2\rangle$ and $\vec{v} = \langle v_1, v_2\rangle$ be vectors in $\mathbb{R}^2$, and let c be a scalar.
   (a) The addition, or sum, of the vectors $\vec{u}$ and $\vec{v}$ is the vector

$$\vec{u} + \vec{v} = \langle u_1 + v_1, u_2 + v_2\rangle. \tag{10.2.7}$$

   (b) The scalar product of c and $\vec{v}$ is the vector

$$c\vec{v} = c\langle v_1, v_2\rangle = \langle cv_1, cv_2\rangle. \tag{10.2.8}$$

2. Let $\vec{u} = \langle u_1, u_2, u_3\rangle$ and $\vec{v} = \langle v_1, v_2, v_3\rangle$ be vectors in $\mathbb{R}^3$, and let c be a scalar.
   (a) The addition, or sum, of the vectors $\vec{u}$ and $\vec{v}$ is the vector

$$\vec{u} + \vec{v} = \langle u_1 + v_1, u_2 + v_2, u_3 + v_3\rangle. \tag{10.2.9}$$

   (b) The scalar product of c and $\vec{v}$ is the vector

$$c\vec{v} = c\langle v_1, v_2, v_3\rangle = \langle cv_1, cv_2, cv_3\rangle. \tag{10.2.10}$$

In short, we say addition and scalar multiplication are computed "component--wise."

Example 10.2.2: Adding vectors

Sketch the vectors $\vec{u} = \langle 1, 3\rangle$, $\vec{v} = \langle 2, 1\rangle$ and $\vec{u} + \vec{v}$ all with initial point at the origin.

Solution

We first compute $\vec{u} + \vec{v}$.

$$\begin{aligned} \vec{u} + \vec{v} &= \langle 1, 3\rangle + \langle 2, 1\rangle \\ &= \langle 3, 4\rangle. \end{aligned}$$

These are all sketched in Figure 10.21.

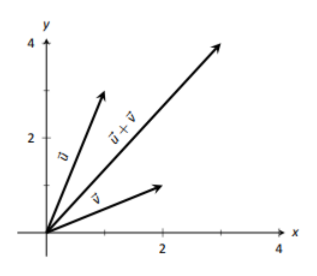


Figure 10.21: Graphing the sum of vectors in Example 10.2.2

As vectors convey magnitude and direction information, the sum of vectors also convey length and magnitude information. Adding $\vec{u}+\vec{v}$ suggests the following idea:

$$\text{"Starting at an initial point, go out } \vec{u}\text{, then go out } \vec{v}\text{."} \tag{10.2.11}$$

This idea is sketched in Figure 10.22, where the initial point of $\vec{v}$ is the terminal point of $\vec{u}$. This is known as the "Head to Tail Rule" of adding vectors. Vector addition is very important. For instance, if the vectors $\vec{u}$ and $\vec{v}$ represent forces acting on a body, the sum $\vec{u}+\vec{v}$ gives the resulting force. Because of various physical applications of vector addition, the sum $\vec{u}+\vec{v}$ is often referred to as the **resultant vector**, or just the "resultant."

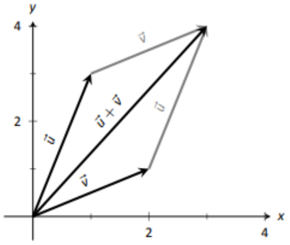


Figure 10.22: Illustrating how to add vectors using the Head to Tail Rule and Parallelogram Law.

Analytically, it is easy to see that $\vec{u}+\vec{v}=\vec{v}+\vec{u}$. Figure 10.22 also gives a graphical representation of this, using gray vectors. Note that the vectors $\vec{u}$ and $\vec{v}$, when arranged as in the figure, form a parallelogram. Because of this, the Head to Tail Rule is also known as the Parallelogram Law: the vector $\vec{u}+\vec{v}$ is defined by forming the parallelogram defined by the vectors $\vec{u}$ and $\vec{v}$; the initial point of $\vec{u}+\vec{v}$ is the common initial point of parallelogram, and the terminal point of the sum is the common terminal point of the parallelogram.

While not illustrated here, the Head to Tail Rule and Parallelogram Law hold for vectors in $\mathbb{R}^3$ as well.

It follows from the properties of the real numbers and Definition 53 that

$$\vec{u}-\vec{v}=\vec{u}+(-1)\vec{v}. \tag{10.2.12}$$

The Parallelogram Law gives us a good way to visualize this subtraction. We demonstrate this in the following example.

Example 10.2.3: Vector Subtraction

Let $\vec{u}=\langle 3,1\rangle$ and $\vec{v}=\langle 1,2\rangle$. Compute and sketch $\vec{u}-\vec{v}$.

Solution

The computation of $\vec{u}-\vec{v}$ is straightforward, and we show all steps below. Usually the formal step of multiplying by (-1) is omitted and we "just subtract."

$$\begin{aligned}\vec{u}-\vec{v}&=\vec{u}+(-1)\vec{v}\\&=\langle 3,1\rangle+\langle -1,-2\rangle\\&=\langle 2,-1\rangle.\end{aligned}$$

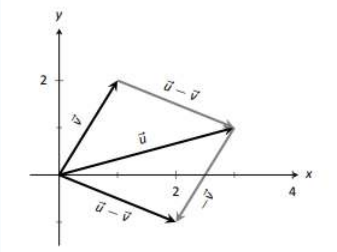


Figure 10.23: Illustrating how to subtract vectors graphically.

Figure 10.23 illustrates, using the Head to Tail Rule, how the subtraction can be viewed as the sum $\vec{u}+(-\vec{v})$. The figure also illustrates how $\vec{u}-\vec{v}$ can be obtained by looking only at the terminal points of $\vec{u}$ and $\vec{v}$ (when their initial points are the same).

Example 10.2.4: Scaling vectors

1. Sketch the vectors $\vec{v}=\langle 2,1\rangle$ and $2\vec{v}$ with initial point at the origin.
2. Compute the magnitudes of $\vec{v}$ and $2\vec{v}$.

Solution

1. We compute $2\vec{v}$:

$$\begin{aligned} 2\vec{v} &= 2\langle 2,1\rangle \\ &= \langle 4,2\rangle . \end{aligned}$$

   Both $\vec{v}$ and $2\vec{v}$ are sketched in Figure10.24. Make note that $2\vec{v}$ does not start at the terminal point of $\vec{v}$; rather, its initial point is also the origin.

2. The figure suggests that $2\vec{v}$ is twice as long as $\vec{v}$. We compute their magnitudes to confirm this.

$$\begin{aligned} \|\vec{v}\| &= \sqrt{2^2+1^2} \\ &= \sqrt{5}. \\ \|2\vec{v}\| &= \sqrt{4^2+2^2} \\ &= \sqrt{20} \\ &= \sqrt{4\cdot 5} = 2\sqrt{5}. \end{aligned}$$

   As we suspected, $2\vec{v}$ is twice as long as $\vec{v}$.

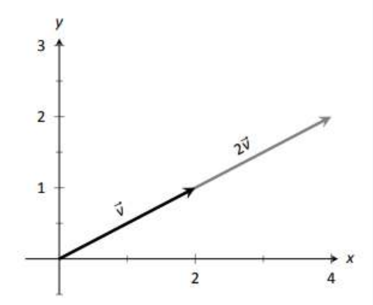


Figure 10.24: Graphing vectors $\vec{v}$ and $2\vec{v}$ in Example 10.2.4

The **zero vector** is the vector whose initial point is also its terminal point. It is denoted by $\vec{0}$. Its component form, in $\mathbb{R}^2$, is $\langle 0,0\rangle$; in $\mathbb{R}^3$, it is $\langle 0,0,0\rangle$. Usually the context makes is clear whether $\vec{0}$ is referring to a vector in the plane or in space.

Our examples have illustrated key principles in vector algebra: how to add and subtract vectors and how to multiply vectors by a scalar. The following theorem states formally the properties of these operations.

THEOREM 84 PROPERTIES OF VECTOR OPERATIONS

The following are true for all scalars c and d, and for all vectors $\vec{u}$, $\vec{v}$ and $\vec{w}$, where $\vec{u}$, $\vec{v}$ and $\vec{w}$ are all in $\mathbb{R}^2$ or where $\vec{u}$, $\vec{v}$ and $\vec{w}$ are all in $\mathbb{R}^3$:

1. $\underbrace{\vec{u}+\vec{v}=\vec{v}+\vec{u}}_{Commutative Property}$
2. $\underbrace{(\vec{u}+\vec{v})+\vec{w}=\vec{u}+(\vec{v}+\vec{w})}_{Associative Property}$
3. $\underbrace{\vec{v}+\vec{0}=\vec{v}}_{Additive Identity}$
4. $(cd)\vec{v}=c(d\vec{v})$
5. $\underbrace{c(\vec{u}+\vec{v})=c\vec{u}+c\vec{v}}_{Distributive Property}$
6. $\underbrace{(c+d)\vec{v}=c\vec{v}+d\vec{v}}_{Distributive Property}$
7. $0\vec{v}=\vec{0}$
8. $\|c\vec{v}\|=|c|\cdot\|\vec{v}\|$
9. $\|u\|=0$ if, and only if, $\vec{u}=\vec{0}$.

As stated before, each vector $\vec{v}$ conveys magnitude and direction information. We have a method of extracting the magnitude, which we write as $\|\vec{v}\|$. *Unit vectors* are a way of extracting just the direction information from a vector.

Definition 54 Unit Vector

A **unit vector** is a vector $\vec{v}$ with a magnitude of 1; that is,

$$\|\vec{v}\|=1. \tag{10.2.13}$$

Consider this scenario: you are given a vector $\vec{v}$ and are told to create a vector of length 10 in the direction of $\vec{v}$. How does one do that? If we knew that $\vec{u}$ was the unit vector in the direction of $\vec{v}$, the answer would be easy: $10\vec{u}$. So how do we find $\vec{u}$?

Property 8 of Theorem 84 holds the key. If we divide $\vec{v}$ by its magnitude, it becomes a vector of length 1. Consider:

$$\begin{aligned}\left\|\frac{1}{\|\vec{v}\|}\vec{v}\right\| &= \frac{1}{\|\vec{v}\|}\|\vec{v}\| \qquad \text{(we can pull out } \frac{1}{\|\vec{v}\|} \text{ as it is a scalar)}\\ &= 1.\end{aligned}$$

So the vector of length 10 in the direction of $\vec{v}$ is $10\frac{1}{\|\vec{v}\|}\vec{v}$. An example will make this more clear.

Example 10.2.5: Using Unit Vectors

Let $\vec{v}=\langle 3,1\rangle$ and let $\vec{w}=\langle 1,2,2\rangle$.

1. Find the unit vector in the direction of $\vec{v}$.
2. Find the unit vector in the direction of $\vec{w}$.
3. Find the vector in the direction of $\vec{v}$ with magnitude 5.

Solution

1. We find $\|\vec{v}\|=\sqrt{10}$. So the unit vector $\vec{u}$ in the direction of $\vec{v}$ is

$$\vec{u}=\frac{1}{\sqrt{10}}\vec{v}=\langle\frac{3}{\sqrt{10}},\frac{1}{\sqrt{10}}\rangle. \tag{10.2.14}$$

2. We find $\|\vec{w}\|=3$, so the unit vector $\vec{z}$ in the direction of $\vec{w}$ is

$$\vec{u}=\frac{1}{3}\vec{w}=\langle\frac{1}{3},\frac{2}{3},\frac{2}{3}\rangle. \tag{10.2.15}$$

3. To create a vector with magnitude 5 in the direction of $\vec{v}$, we multiply the unit vector $\vec{u}$ by 5. Thus $5\vec{u}=\langle 15/\sqrt{10},5/\sqrt{10}\rangle$ is the vector we seek. This is sketched in Figure 10.25.

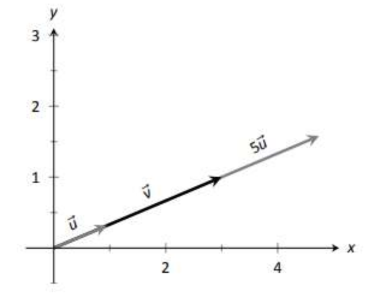


Figure 10.25: Graphing vectors in Example 10.2.5. All vectors shown have their initial points at the origin.

The basic formation of the unit vector $\vec u$ in the direction of a vector $\vec v$ leads to a interesting equation. It is:

$$\vec v = \|\vec v\| \frac{1}{\|\vec v\|} \vec v. \tag{10.2.16}$$

We rewrite the equation with parentheses to make a point:

$$\vec v = \underbrace{\|\vec v\|}_{\text{magnitude}} \cdot \underbrace{\left(\frac{1}{\|\vec v\|}\vec v\right)}_{\text{direction}}.$$

This equation illustrates the fact that a vector has both magnitude and direction, where we view a unit vector as supplying *only* direction information. Identifying unit vectors with direction allows us to define **parallel vectors**.

Definition 55 Parallel Vectors

1. Unit vectors $\vec u_1$ and $\vec u_2$ are **parallel** if $\vec u_1 = \pm \vec u_2$.
2. Nonzero vectors $\vec v_1$ and $\vec v_2$ are **parallel** if their respective unit vectors are parallel.

It is equivalent to say that vectors $\vec v_1$ and $\vec v_2$ are parallel if there is a scalar $c \neq 0$ such that $\vec v_1 = c\vec v_2$ (see marginal note).

Note: $\vec 0$ is directionless; because $\|0\| = 0$, there is no unit vector in the "direction" of $\vec 0$.

Some texts define two vectors as being parallel if one is a scalar multiple of the other. By this definition, $\vec 0$ is parallel to all vectors as $\vec 0 = 0\vec v$ for all $\vec v$.

We prefer the given definition of parallel as it is grounded in the fact that unit vectors provide direction information. One may adopt the convention that $\vec 0$ is parallel to all vectors if they desire.

If one graphed all unit vectors in $\mathbb{R}^2$ with the initial point at the origin, then the terminal points would all lie on the unit circle. Based on what we know from trigonometry, we can then say that the component form of all unit vectors in $\mathbb{R}^2$ is $\langle \cos\theta, \sin\theta \rangle$ for some angle θ.

A similar construction in $\mathbb{R}^3$ shows that the terminal points all lie on the unit sphere. These vectors also have a particular component form, but its derivation is not as straightforward as the one for unit vectors in $\mathbb{R}^2$. Important concepts about unit vectors are given in the following Key Idea.

KEY IDEA 48 UNIT VECTORS

1. The unit vector in the direction of $\vec v$ is

$$\vec u = \frac{1}{\|\vec v\|} \vec v. \tag{10.2.17}$$

2. A vector $\vec u$ in $\mathbb{R}^2$ is a unit vector if, and only if, its component form is $\langle \cos\theta, \sin\theta \rangle$ for some angle θ.
3. A vector $\vec u$ in $\mathbb{R}^3$ is a unit vector if, and only if, its component form is $\langle \sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta \rangle$ for some angles θ and φ.

These formulas can come in handy in a variety of situations, especially the formula for unit vectors in the plane.

Example 10.2.6: Finding Component Forces

Consider a weight of 50lb hanging from two chains, as shown in Figure 10.26. One chain makes an angle of 30° with the vertical, and the other an angle of 45°. Find the force applied to each chain.

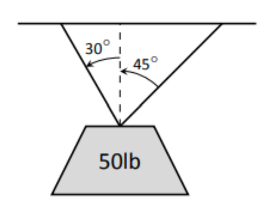


Figure 10.26: A diagram of a weight hanging from 2 chains in Example 10.2.6.

Solution

Knowing that gravity is pulling the 50lb weight straight down, we can create a vector $\vec{F}$ to represent this force.

$$\vec{F}=50\langle 0,-1\rangle=\langle 0,-50\rangle. \tag{10.2.18}$$

We can view each chain as "pulling" the weight up, preventing it from falling. We can represent the force from each chain with a vector. Let $\vec{F}_1$ represent the force from the chain making an angle of 30° with the vertical, and let $\vec{F}_2$ represent the force form the other chain. Convert all angles to be measured from the horizontal (as shown in Figure 10.27), and apply Key Idea 48. As we do not yet know the magnitudes of these vectors, (that is the problem at hand), we use m_1 and m_2 to represent them.

$$\vec{F}_1=m_1\langle\cos 120^\circ,\sin 120^\circ\rangle \tag{10.2.19}$$

$$\vec{F}_2=m_2\langle\cos 45^\circ,\sin 45^\circ\rangle \tag{10.2.20}$$

As the weight is not moving, we know the sum of the forces is $\vec{0}$. This gives:

$$\begin{aligned}\vec{F}+\vec{F}_1+\vec{F}_2&=\vec{0}\\ \langle 0,-50\rangle+m_1\langle\cos 120^\circ,\sin 120^\circ\rangle+m_2\langle\cos 45^\circ,\sin 45^\circ\rangle&=\vec{0}\end{aligned}$$

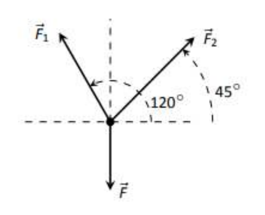


Figure 10.27: A diagram of the force vectors from Example 10.2.6.

The sum of the entries in the first component is 0, and the sum of the entries in the second component is also 0. This leads us to the following two equations:

$$\begin{aligned}m_1\cos 120^\circ+m_2\cos 45^\circ&=0\\ m_1\sin 120^\circ+m_2\sin 45^\circ&=50\end{aligned}$$

This is a simple 2-equation, 2-unkown system of linear equations. We leave it to the reader to verify that the solution is

$$m_1=50(\sqrt{3}-1)\approx 36.6;\qquad m_2=\frac{50\sqrt{2}}{1+\sqrt{3}}\approx 25.88. \tag{10.2.21}$$

It might seem odd that the sum of the forces applied to the chains is more than 50lb. We leave it to a physics class to discuss the full details, but offer this short explanation. Our equations were established so that the *vertical* components of each force sums to 50lb, thus supporting the weight. Since the chains are at an angle, they also pull against each other, creating an "additional" horizontal force while holding the weight in place.

Unit vectors were very important in the previous calculation; they allowed us to define a vector in the proper direction but with an unknown magnitude. Our computations were then computed component--wise. Because such calculations are often necessary, the *standard unit vectors* can be useful.

Definition 56 STANDARD UNIT VECTORS

1. In $\mathbb{R}^2$, the standard unit vectors are

$$\vec{i} = \langle 1, 0\rangle \quad \text{and} \quad \vec{j} = \langle 0, 1\rangle. \tag{10.2.22}$$

2. In $\mathbb{R}^3$, the standard unit vectors are

$$\vec{i} = \langle 1, 0, 0\rangle \quad \text{and} \quad \vec{j} = \langle 0, 1, 0\rangle \quad \text{and} \quad \vec{k} = \langle 0, 0, 1\rangle. \tag{10.2.23}$$

Example 10.2.7: Using standard unit vectors

1. Rewrite $\vec{v} = \langle 2, -3\rangle$ using the standard unit vectors.
2. Rewrite $\vec{w} = 4\vec{i} - 5\vec{j} + 2\vec{k}$ in component form.

Solution

1.

$$\begin{aligned} \vec{v} &= \langle 2, -3\rangle \tag{10.2.24}\\ &= \langle 2, 0\rangle + \langle 0, -3\rangle \tag{10.2.25}\\ &= 2\langle 1, 0\rangle - 3\langle 0, 1\rangle \tag{10.2.26}\\ &= 2\vec{i} - 3\vec{j} \tag{10.2.27}\end{aligned}$$

2.

$$\begin{aligned} \vec{w} &= 4\vec{i} - 5\vec{j} + 2\vec{k} \tag{10.2.28}\\ &= \langle 4, 0, 0\rangle + \langle 0, -5, 0\rangle + \langle 0, 0, 2\rangle \tag{10.2.29}\\ &= \langle 4, -5, 2\rangle \tag{10.2.30}\end{aligned}$$

These two examples demonstrate that converting between component form and the standard unit vectors is rather straightforward. Many mathematicians prefer component form, and it is the preferred notation in this text. Many engineers prefer using the standard unit vectors, and many engineering text use that notation.

Example 10.2.8: Finding Component Force

A weight of 25lb is suspended from a chain of length 2ft while a wind pushes the weight to the right with constant force of 5lb as shown in Figure 10.28. What angle will the chain make with the vertical as a result of the wind's pushing? How much higher will the weight be?

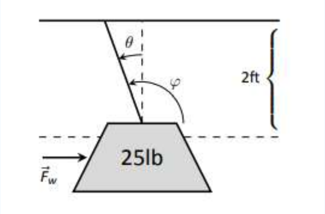


Figure 10.28: A figure of a weight being pushed by the wind in Example 10.2.8.

Solution

The force of the wind is represented by the vector $\vec{F}_w = 5\vec{i}$. The force of gravity on the weight is represented by $\vec{F}_g = -25\vec{j}$. The direction and magnitude of the vector representing the force on the chain are both unknown. We represent this force with

$$\vec{F}_c = m\langle \cos\varphi, \sin\varphi\rangle = m\cos\varphi\,\vec{i} + m\sin\varphi\,\vec{j} \tag{10.2.31}$$

for some magnitude m and some angle with the horizontal φ. (Note: θ is the angle the chain makes with the *vertical*; φ is the angle with the *horizontal*.)

As the weight is at equilibrium, the sum of the forces is $\vec{0}$:

$$\begin{aligned} \vec{F}_c + \vec{F}_w + \vec{F}_g &= \vec{0}\\ m\cos\varphi\,\vec{i} + m\sin\varphi\,\vec{j} + 5\vec{i} - 25\vec{j} &= \vec{0}\end{aligned}$$

Thus the sum of the $\vec{i}$ and $\vec{j}$ components are 0, leading us to the following system of equations:

$$5+m\cos\varphi=0 \tag{10.2.32}$$

$$-25+m\sin\varphi=0 \tag{10.2.33}$$

This is enough to determine $\vec{F}_c$ already, as we know $m\cos\varphi=-5$ and $m\sin\varphi=25$. Thus $F_c=\langle -5,25\rangle$. We can use this to find the magnitude m:

$$m=\sqrt{(-5)^2+25^2}=5\sqrt{26}\approx 25.5\text{lb}. \tag{10.2.34}$$

We can then use either equality from Equation 10.2.33 to solve for φ. We choose the first equality as using arccosine will return an angle in the 2^{nd} quadrant:

$$5+5\sqrt{26}\cos\varphi=0 \quad\Rightarrow\quad \varphi=\cos^{-1}\left(\frac{-5}{5\sqrt{26}}\right)\approx 1.7682\approx 101.31^{\circ}. \tag{10.2.35}$$

Subtracting 90° from this angle gives us an angle of 11.31° with the vertical.

We can now use trigonometry to find out how high the weight is lifted. The diagram shows that a right triangle is formed with the 2ft chain as the hypotenuse with an interior angle of 11.31°. The length of the adjacent side (in the diagram, the dashed vertical line) is $2\cos 11.31^{\circ}\approx 1.96$ft. Thus the weight is lifted by about 0.04ft, almost 1/2in.

The algebra we have applied to vectors is already demonstrating itself to be very useful. There are two more fundamental operations we can perform with vectors, the *dot product* and the cross product. The next two sections explore each in turn.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 10.2: An Introduction to Vectors is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

10.3: The Dot Product

The previous section introduced vectors and described how to add them together and how to multiply them by scalars. This section introduces *a* multiplication on vectors called the **dot product**.

Definition 57 Dot Product

1. Let $\vec{u} = \langle u_1, u_2\rangle$ and $\vec{v} = \langle v_1, v_2\rangle$ in $\mathbb{R}^2$. The **dot product** of $\vec{u}$ and $\vec{v}$, denoted $\vec{u}\cdot\vec{v}$, is

$$\vec{u}\cdot\vec{v} = u_1v_1 + u_2v_2. \tag{10.3.1}$$

2. Let $\vec{u} = \langle u_1, u_2, u_3\rangle$ and $\vec{v} = \langle v_1, v_2, v_3\rangle$ in $\mathbb{R}^3$. The **dot product** of $\vec{u}$ and $\vec{v}$, denoted $\vec{u}\cdot\vec{v}$, is

$$\vec{u}\cdot\vec{v} = u_1v_1 + u_2v_2 + u_3v_3. \tag{10.3.2}$$

Note how this product of vectors returns a *scalar*, not another vector. We practice evaluating a dot product in the following example, then we will discuss why this product is useful.

Example $10.3.1$: Evaluating dot products

1. Let $\vec{u} = \langle 1, 2\rangle$, $\vec{v} = \langle 3, -1\rangle$ in $\mathbb{R}^2$. Find $\vec{u}\cdot\vec{v}$.
2. Let $\vec{x} = \langle 2, -2, 5\rangle$ and $\vec{y} = \langle -1, 0, 3\rangle$ in $\mathbb{R}^3$. Find $\vec{x}\cdot\vec{y}$.

Solution

1. Using Definition 57, we have

$$\vec{u}\cdot\vec{v} = 1(3) + 2(-1) = 1. \tag{10.3.3}$$

2. Using the definition, we have

$$\vec{x}\cdot\vec{y} = 2(-1) - 2(0) + 5(3) = 13. \tag{10.3.4}$$

The dot product, as shown by the preceding example, is very simple to evaluate. It is only the sum of products. While the definition gives no hint as to why we would care about this operation, there is an amazing connection between the dot product and angles formed by the vectors. Before stating this connection, we give a theorem stating some of the properties of the dot product.

THEOREM 85 PROPERTIES OF THE DOT PRODUCT

Let $\vec{u}$, $\vec{v}$ and $\vec{w}$ be vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$ and let c be a scalar.

1. $\vec{u}\cdot\vec{v} = \vec{v}\cdot\vec{u}$ Commutative Property
2. $\vec{u}\cdot(\vec{v}+\vec{w}) = \vec{u}\cdot\vec{v} + \vec{u}\cdot\vec{w}$ Distributive Property
3. $c(\vec{u}\cdot\vec{v}) = (c\vec{u})\cdot\vec{v} = \vec{u}\cdot(c\vec{v})$
4. $\vec{0}\cdot\vec{v} = 0$
5. $\vec{v}\cdot\vec{v} = \|\vec{v}\|^2$

The last statement of the theorem makes a handy connection between the magnitude of a vector and the dot product with itself. Our definition and theorem give properties of the dot product, but we are still likely wondering "What does the dot product *mean*?" It is helpful to understand that the dot product of a vector with itself is connected to its magnitude.

The next theorem extends this understanding by connecting the dot product to magnitudes and angles. Given vectors $\vec{u}$ and $\vec{v}$ in the plane, an angle θ is clearly formed when $\vec{u}$ and $\vec{v}$ are drawn with the same initial point as illustrated in Figure 10.29(a). (We always take θ to be the angle in $[0, \pi]$ as two angles are actually created.)

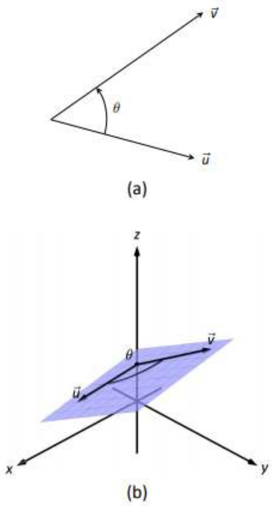


Figure 10.29: Illustrating the angle formed by two vectors with the same initial point.

The same is also true of 2 vectors in space: given $\vec{u}$ and $\vec{v}$ in $\mathbb{R}^3$ with the same initial point, there is a plane that contains both $\vec{u}$ and $\vec{v}$. (When $\vec{u}$ and $\vec{v}$ are co-linear, there are infinite planes that contain both vectors.) In that plane, we can again find an angle θ between them (and again, $0 \leq \theta \leq \pi$). This is illustrated in Figure 10.29(b).

The following theorem connects this angle θ to the dot product of $\vec{u}$ and $\vec{v}$.

theorem 86 the dot product and angles

Let $\vec{u}$ and $\vec{v}$ be vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$. Then

$$\vec{u} \cdot \vec{v} = \|\vec{u}\|\,\|\vec{v}\| \cos\theta, \tag{10.3.5}$$

where θ, $0 \leq \theta \leq \pi$, is the angle between $\vec{u}$ and $\vec{v}$.

When θ is an acute angle (i.e., $0 \leq \theta < \pi/2$), $\cos\theta$ is positive; when $\theta = \pi/2$, $\cos\theta = 0$; when θ is an obtuse angle ($\pi/2 < \theta \leq \pi$), $\cos\theta$ is negative. Thus the sign of the dot product gives a general indication of the angle between the vectors, illustrated in Figure 10.30.

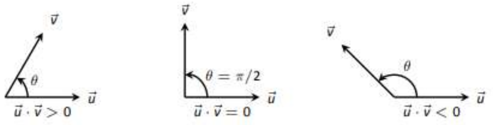


Figure 10.30: Illustrating the relationship between the angle between vectors and the sign of their dot product.

We *can* use Theorem 86 to compute the dot product, but generally this theorem is used to find the angle between known vectors (since the dot product is generally easy to compute). To this end, we rewrite the theorem's equation as

$$\cos\theta = \frac{\vec{u}\cdot\vec{v}}{\|\vec{u}\|\|\vec{v}\|} \quad \Leftrightarrow \quad \theta = \cos^{-1}\left(\frac{\vec{u}\cdot\vec{v}}{\|\vec{u}\|\|\vec{v}\|}\right). \tag{10.3.6}$$

We practice using this theorem in the following example.

Example 10.3.2: Using the dot product to find angles

Let $\vec{u} = \langle 3, 1\rangle$, $\vec{v} = \langle -2, 6\rangle$ and $\vec{w} = \langle -4, 3\rangle$, as shown in Figure 10.31. Find the angles α, β and θ.

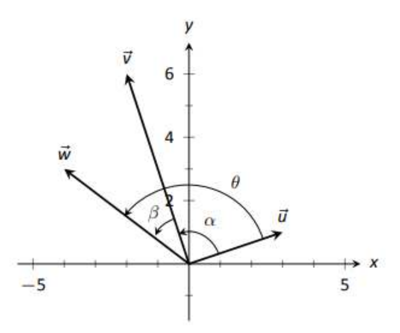


Figure 10.31: Vectors used in Example 10.3.2.

Solution

We start by computing the magnitude of each vector.

$$\|\vec{u}\| = \sqrt{10}; \quad \|\vec{v}\| = 2\sqrt{10}; \quad \|\vec{w}\| = 5. \tag{10.3.7}$$

We now apply Theorem 86 to find the angles.

$$\begin{aligned}\alpha &= \cos^{-1}\left(\frac{\vec{u}\cdot\vec{v}}{(\sqrt{10})(2\sqrt{10})}\right)\\ &= \cos^{-1}(0) = \frac{\pi}{2} = 90^\circ.\\ \beta &= \cos^{-1}\left(\frac{\vec{v}\cdot\vec{w}}{(2\sqrt{10})(5)}\right)\\ &= \cos^{-1}\left(\frac{26}{10\sqrt{10}}\right)\\ &\approx 0.6055 \approx 34.7^\circ.\\ \theta &= \cos^{-1}\left(\frac{\vec{u}\cdot\vec{w}}{(\sqrt{10})(5)}\right)\\ &= \cos^{-1}\left(\frac{-9}{5\sqrt{10}}\right)\\ &\approx 2.1763 \approx 124.7^\circ\end{aligned}$$

We see from our computation that $\alpha+\beta=\theta$, as indicated by Figure 10.31. While we knew this should be the case, it is nice to see that this non-intuitive formula indeed returns the results we expected.

We do a similar example next in the context of vectors in space.

Example $10.3.3$: Using the dot product to find angles

Let $\vec{u}=\langle 1,1,1\rangle$, $\vec{v}=\langle -1,3,-2\rangle$ and $\vec{w}=\langle -5,1,4\rangle$, as illustrated in Figure 10.32. Find the angle between each pair of vectors.

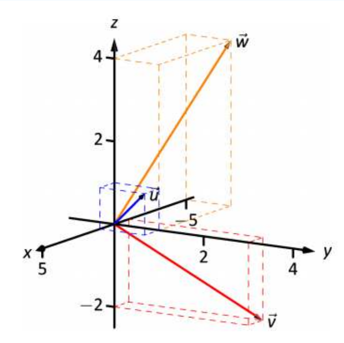


Figure 10.32: Vectors used in Example 10.3.3

Solution

1. Between $\vec{u}$ and $\vec{v}$:

$$\begin{aligned}\theta &= \cos^{-1}\left(\frac{\vec{u}\cdot\vec{v}}{\|\vec{u}\|\|\vec{v}\|}\right)\\ &= \cos^{-1}\left(\frac{0}{\sqrt{3}\sqrt{14}}\right)\\ &= \frac{\pi}{2}.\end{aligned}$$

2. Between $\vec{u}$ and $\vec{w}$:

$$\begin{aligned}\theta &= \cos^{-1}\left(\frac{\vec{u}\cdot\vec{w}}{\|\vec{u}\|\|\vec{w}\|}\right)\\ &= \cos^{-1}\left(\frac{0}{\sqrt{3}\sqrt{42}}\right)\\ &= \frac{\pi}{2}.\end{aligned}$$

3. Between $\vec{v}$ and $\vec{w}$:

$$\begin{aligned}\theta &= \cos^{-1}\left(\frac{\vec{v}\cdot\vec{w}}{\|\vec{v}\|\|\vec{w}\|}\right)\\ &= \cos^{-1}\left(\frac{0}{\sqrt{14}\sqrt{42}}\right)\\ &= \frac{\pi}{2}.\end{aligned}$$

While our work shows that each angle is $\pi/2$, i.e., 90°, none of these angles looks to be a right angle in Figure 10.32. Such is the case when drawing three--dimensional objects on the page.

All three angles between these vectors was $\pi/2$, or 90°. We know from geometry and everyday life that 90° angles are "nice" for a variety of reasons, so it should seem significant that these angles are all $\pi/2$. Notice the common feature in each calculation (and also the calculation of α in Example 10.3.2): the dot products of each pair of angles was 0. We use this as a basis for a definition of the term **orthogonal**, which is essentially synonymous to *perpendicular*.

Definition 58 ORTHOGONAL

Vectors $\vec{u}$ and $\vec{v}$ are **orthogonal** if their dot product is 0.

Note: The term *perpendicular* originally referred to lines. As mathematics progressed, the concept of "being at right angles to" was applied to other objects, such as vectors and planes, and the term *orthogonal* was introduced. It is especially used when discussing objects that are hard, or impossible, to visualize: two vectors in 5-dimensional space are orthogonal if their dot product is 0. It is not wrong to say they are *perpendicular*, but common convention gives preference to the word *orthogonal*.

Example 10.3.4: Finding orthogonal vectors

Let $\vec{u}=\langle 3,5\rangle$ and $\vec{v}=\langle 1,2,3\rangle$.

1. Find two vectors in $\mathbb{R}^2$ that are orthogonal to $\vec{u}$.
2. Find two non--parallel vectors in $\mathbb{R}^3$ that are orthogonal to $\vec{v}$.

Solution

1. Recall that a line perpendicular to a line with slope m has slope $-1/m$, the "opposite reciprocal slope." We can think of the slope of $\vec{u}$ as $5/3$, its "rise over run." A vector orthogonal to $\vec{u}$ will have slope $-3/5$. There are many such choices, though all parallel:

$$\langle -5,3\rangle \quad \text{or} \quad \langle 5,-3\rangle \quad \text{or} \quad \langle -10,6\rangle \quad \text{or} \quad \langle 15,-9\rangle, \text{etc.} \tag{10.3.8}$$

2. There are infinite directions in space orthogonal to any given direction, so there are an infinite number of non--parallel vectors orthogonal to $\vec{v}$. Since there are so many, we have great leeway in finding some.

   One way is to arbitrarily pick values for the first two components, leaving the third unknown. For instance, let $\vec{v}_1 = \langle 2, 7, z\rangle$. If $\vec{v}_1$ is to be orthogonal to $\vec{v}$, then $\vec{v}_1 \cdot \vec{v} = 0$, so

   $$2 + 14 + 3z = 0 \quad \Rightarrow z = \frac{-16}{3}. \tag{10.3.9}$$

   So $\vec{v}_1 = \langle 2, 7, -16/3\rangle$ is orthogonal to $\vec{v}$. We can apply a similar technique by leaving the first or second component unknown.

   Another method of finding a vector orthogonal to $\vec{v}$ mirrors what we did in part 1. Let $\vec{v}_2 = \langle -2, 1, 0\rangle$. Here we switched the first two components of $\vec{v}$, changing the sign of one of them (similar to the "opposite reciprocal" concept before). Letting the third component be 0 effectively ignores the third component of $\vec{v}$, and it is easy to see that

   $$\vec{v}_2 \cdot \vec{v} = \langle -2, 1, 0\rangle \cdot \langle 1, 2, 3\rangle = 0. \tag{10.3.10}$$

   Clearly $\vec{v}_1$ and $\vec{v}_2$ are not parallel.

An important construction is illustrated in Figure 10.33, where vectors $\vec{u}$ and $\vec{v}$ are sketched. In part (a), a dotted line is drawn from the tip of $\vec{u}$ to the line containing $\vec{v}$, where the dotted line is orthogonal to $\vec{v}$. In part (b), the dotted line is replaced with the vector $\vec{z}$ and $\vec{w}$ is formed, parallel to $\vec{v}$. It is clear by the diagram that $\vec{u} = \vec{w} + \vec{z}$. What is important about this construction is this: $\vec{u}$ is *decomposed* as the sum of two vectors, one of which is parallel to $\vec{v}$ and one that is perpendicular to $\vec{v}$. It is hard to overstate the importance of this construction (as we'll see in upcoming examples).

The vectors $\vec{w}$, $\vec{z}$ and $\vec{u}$ as shown in Figure 10.33 (b) form a right triangle, where the angle between $\vec{v}$ and $\vec{u}$ is labeled θ. We can find $\vec{w}$ in terms of $\vec{v}$ and $\vec{u}$.

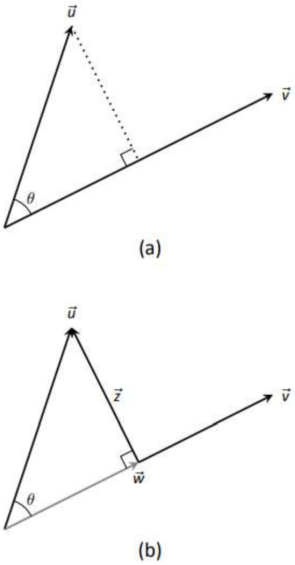


Figure 10.33: Developing the construction of the orthogonal projection.

Using trigonometry, we can state that

$$\|\vec{w}\| = \|\vec{u}\| \cos\theta. \tag{10.3.11}$$

We also know that $\vec{w}$ is parallel to to $\vec{v}$; that is, the direction of $\vec{w}$ is the direction of $\vec{v}$, described by the unit vector $\frac{1}{\|\vec{v}\|}\vec{v}$. The vector $\vec{w}$ is the vector in the direction $\frac{1}{\|\vec{v}\|}\vec{v}$ with magnitude $\|\vec{u}\| \cos\theta$:

$$\vec{w} = \Big(\|\vec{u}\|\cos\theta\Big)\frac{1}{\|\vec{v}\|}\vec{v}.$$

Replace $\cos\theta$ using Theorem 86:

$$= \left(\|\vec{u}\|\frac{\vec{u}\cdot\vec{v}}{\|\vec{u}\|\|\vec{v}\|}\right)\frac{1}{\|\vec{v}\|}\vec{v}$$

$$= \frac{\vec{u}\cdot\vec{v}}{\|\vec{v}\|^2}\vec{v}.$$

Now apply Theorem 85.

$$= \frac{\vec{u}\cdot\vec{v}}{\vec{v}\cdot\vec{v}}\vec{v}.$$

Since this construction is so important, it is given a special name.

Definition 59 orthogonal projection

Let $\vec{u}$ and $\vec{v}$ be given. The **orthogonal projection of $\vec{u}$ onto $\vec{v}$**, denoted $\text{proj}_{\vec{v}}\vec{u}$, is

$$\text{proj}_{\vec{v}}\vec{u} = \frac{\vec{u}\cdot\vec{v}}{\vec{v}\cdot\vec{v}}\vec{v}. \tag{10.3.12}$$

Example 10.3.5: Computing the orthogonal projection

1. Let $\vec{u} = \langle -2, 1\rangle$ and $\vec{v} = \langle 3, 1\rangle$. Find $\text{proj}_{\vec{v}}\vec{u}$, and sketch all three vectors with initial points at the origin.
2. Let $\vec{w} = \langle 2, 1, 3\rangle$ and $\vec{x} = \langle 1, 1, 1\rangle$. Find $\text{proj}_{\vec{x}}\vec{w}$, and sketch all three vectors with initial points at the origin.

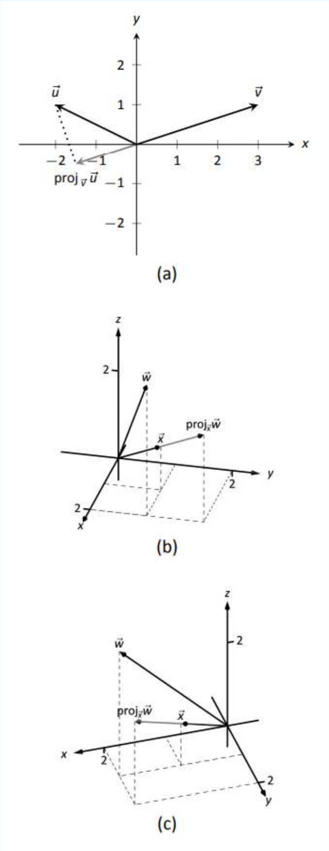

y
2
1
$\vec{u}$
$\vec{v}$
x
−2
−1
1
2
3
$\text{proj}_{\vec{v}}\vec{u}$
−1
−2
(a)
z
2
$\vec{w}$
$\text{proj}_{\vec{x}}\vec{w}$
$\vec{x}$
y
2
2
x
(b)
z
2
$\vec{w}$
$\text{proj}_{\vec{x}}\vec{w}$
$\vec{x}$
x
2
2
y
(c)

Figure 10.34: Graphing the vectors used in Example 10.3.5

Solution

1. Applying Definition 59, we have

$$\begin{aligned}\text{proj}_{\vec{v}}\vec{u} &= \frac{\vec{u}\cdot\vec{v}}{\vec{v}\cdot\vec{v}}\vec{v}\\ &= \frac{-5}{10}\langle 3,1\rangle\\ &= \langle -\frac{3}{2}, -\frac{1}{2}\rangle.\end{aligned}$$

Vectors $\vec{u}$, $\vec{v}$ and $\text{proj}_{\vec{v}}\vec{u}$ are sketched in Figure 10.34(a). Note how the projection is parallel to $\vec{v}$; that is, it lies on the same line through the origin as $\vec{v}$, although it points in the opposite direction. That is because the angle between $\vec{u}$ and $\vec{v}$ is obtuse (i.e., greater than 90°).

2. Apply the definition:

$$\begin{aligned}\text{proj}_{\vec{x}}\vec{w} &= \frac{\vec{w}\cdot\vec{x}}{\vec{x}\cdot\vec{x}}\vec{x}\\ &= \frac{6}{3}\langle 1,1,1\rangle\\ &= \langle 2,2,2\rangle.\end{aligned}$$

These vectors are sketched in Figure 10.34(b), and again in part (c) from a different perspective. Because of the nature of graphing these vectors, the sketch in part (b) makes it difficult to recognize that the drawn projection has the geometric properties it should. The graph shown in part (c) illustrates these properties better.

Consider Figure 10.35 where the concept of the orthogonal projection is again illustrated. It is clear that

$$\vec{u} = \text{proj}_{\vec{v}}\vec{u} + \vec{z}. \tag{10.3.13}$$

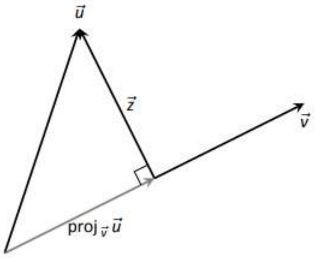


Figure 10.35: Illustrating the orthogonal projection.

As we know what $\vec{u}$ and $\text{proj}_{\vec{v}}\vec{u}$ are, we can solve for $\vec{z}$ and state that

$$\vec{z} = \vec{u} - \text{proj}_{\vec{v}}\vec{u}. \tag{10.3.14}$$

This leads us to rewrite Equation 10.3.13 in a seemingly silly way:

$$\vec{u} = \text{proj}_{\vec{v}}\vec{u} + (\vec{u} - \text{proj}_{\vec{v}}\vec{u}). \tag{10.3.15}$$

This is not nonsense, as pointed out in the following Key Idea. (Notation note: the expression "$\parallel \vec{y}$" means "is parallel to $\vec{y}$." We can use this notation to state "$\vec{x} \parallel \vec{y}$" which means "$\vec{x}$ is parallel to $\vec{y}$." The expression "$\perp \vec{y}$" means "is orthogonal to $\vec{y}$," and is used similarly.)

key idea 49 orthogonal decomposition of vectors

Let $\vec{u}$ and $\vec{v}$ be given. Then $\vec{u}$ can be written as the sum of two vectors, one of which is parallel to $\vec{v}$, and one of which is orthogonal to $\vec{v}$:

$$\vec{u} = \underbrace{\text{proj}_{\vec{v}}\vec{u}}_{\parallel\,\vec{v}} + (\underbrace{\vec{u} - \text{proj}_{\vec{v}}\vec{u}}_{\perp\,\vec{v}}). \tag{10.3.16}$$

We illustrate the use of this equality in the following example.

Example 10.3.6: Orthogonal decomposition of vectors

1. Let $\vec{u} = \langle -2, 1\rangle$ and $\vec{v} = \langle 3, 1\rangle$ as in Example 10.3.5. Decompose $\vec{u}$ as the sum of a vector parallel to $\vec{v}$ and a vector orthogonal to $\vec{v}$.
2. Let $\vec{w} = \langle 2, 1, 3\rangle$ and $\vec{x} = \langle 1, 1, 1\rangle$ as in Example 10.3.5. Decompose $\vec{w}$ as the sum of a vector parallel to $\vec{x}$ and a vector orthogonal to $\vec{x}$.

Solution

1. In Example 10.3.5, we found that $\text{proj}_{\vec{v}}\vec{u} = \langle -1.5, -0.5\rangle$. Let
$$\vec{z} = \vec{u} - \text{proj}_{\vec{v}}\vec{u} = \langle -2, 1\rangle - \langle -1.5, -0.5\rangle = \langle -0.5, 1.5\rangle. \tag{10.3.17}$$
Is $\vec{z}$ orthogonal to $\vec{v}$? (I.e, is $\vec{z} \perp \vec{v}$?) We check for orthogonality with the dot product:
$$\vec{z} \cdot \vec{v} = \langle -0.5, 1.5\rangle \cdot \langle 3, 1\rangle = 0. \tag{10.3.18}$$
Since the dot product is 0, we know $\vec{z} \perp \vec{v}$. Thus:
$$\begin{aligned} \vec{u} &= \text{proj}_{\vec{v}}\vec{u} \;+\; (\vec{u} - \text{proj}_{\vec{v}}\vec{u}) \\ \langle -2, 1\rangle &= \underbrace{\langle -1.5, -0.5\rangle}_{\parallel\, \vec{v}} \;+\; \underbrace{\langle -0.5, 1.5\rangle}_{\perp\, \vec{v}}. \end{aligned}$$
2. We found in Example 10.3.5 that $\text{proj}_{\vec{x}}\vec{w} = \langle 2, 2, 2\rangle$. Applying the Key Idea, we have:
$$\vec{z} = \vec{w} - \text{proj}_{\vec{x}}\vec{w} = \langle 2, 1, 3\rangle - \langle 2, 2, 2\rangle = \langle 0, -1, 1\rangle. \tag{10.3.19}$$
We check to see if $\vec{z} \perp \vec{x}$:
$$\vec{z} \cdot \vec{x} = \langle 0, -1, 1\rangle \cdot \langle 1, 1, 1\rangle = 0. \tag{10.3.20}$$
Since the dot product is 0, we know the two vectors are orthogonal.

We now write $\vec{w}$ as the sum of two vectors, one parallel and one orthogonal to $\vec{x}$:
$$\begin{aligned} \vec{w} &= \text{proj}_{\vec{x}}\vec{w} \;+\; (\vec{w} - \text{proj}_{\vec{x}}\vec{w}) \\ \langle 2, 1, 3\rangle &= \underbrace{\langle 2, 2, 2\rangle}_{\parallel\, \vec{x}} \;+\; \underbrace{\langle 0, -1, 1\rangle}_{\perp\, \vec{x}} \end{aligned}$$

We give an example of where this decomposition is useful.

Example 10.3.7: Orthogonally decomposing a force vector

Consider Figure 10.36(a), showing a box weighing 50lb on a ramp that rises 5ft over a span of 20ft. Find the components of force, and their magnitudes, acting on the box (as sketched in part (b) of the figure):

1. in the direction of the ramp, and
2. orthogonal to the ramp.

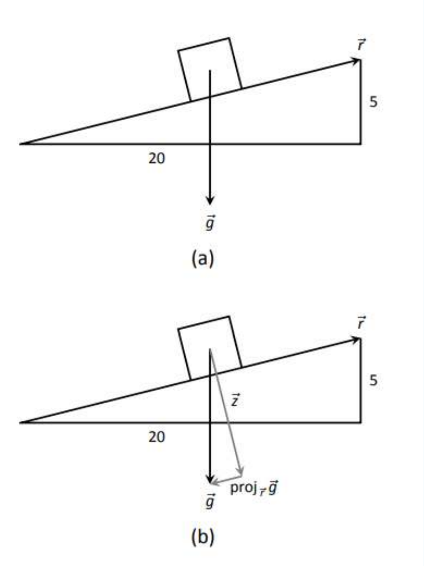


Figure 10.36: Sketching the ramp and box in Example 10.3.7. Note: The vectors are not drawn to scale.

Solution

As the ramp rises 5ft over a horizontal distance of 20ft, we can represent the direction of the ramp with the vector $\vec{r} = \langle 20, 5\rangle$. Gravity pulls down with a force of 50lb, which we represent with $\vec{g} = \langle 0, -50\rangle$.

1. To find the force of gravity in the direction of the ramp, we compute $\text{proj}_{\vec{r}}\vec{g}$:

$$\begin{aligned}\text{proj}_{\vec{r}}\vec{g} &= \frac{\vec{g}\cdot\vec{r}}{\vec{r}\cdot\vec{r}}\vec{r} \\ &= \frac{-250}{425}\langle 20, 5\rangle \\ &= \langle -\frac{200}{17}, -\frac{50}{17}\rangle \approx \langle -11.76, -2.94\rangle.\end{aligned}$$

   The magnitude of $\text{proj}_{\vec{r}}\vec{g}$ is $\|\text{proj}_{\vec{r}}\vec{g}\| = 50/\sqrt{17} \approx 12.13\text{lb}$. Though the box weighs 50lb, a force of about 12lb is enough to keep the box from sliding down the ramp.
2. To find the component $\vec{z}$ of gravity orthogonal to the ramp, we use Key Idea 49.

$$\begin{aligned}\vec{z} &= \vec{g} - \text{proj}_{\vec{r}}\vec{g} \\ &= \langle \frac{200}{17}, -\frac{800}{17}\rangle \approx \langle 11.76, -47.06\rangle.\end{aligned}$$

   The magnitude of this force is $\|\vec{z}\| \approx 48.51$lb. In physics and engineering, knowing this force is important when computing things like static frictional force. (For instance, we could easily compute if the static frictional force alone was enough to keep the box from sliding down the ramp.)

Application to Work

In physics, the application of a force F to move an object in a straight line a distance d produces *work*; the amount of work W is $W = Fd$, (where F is in the direction of travel). The orthogonal projection allows us to compute work when the force is not in the direction of travel.

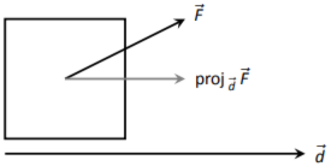


Figure 10.37: Finding work when the force and direction of travel are given as vectors.

Consider Figure 10.37, where a force $\vec{F}$ is being applied to an object moving in the direction of $\vec{d}$. (The distance the object travels is the magnitude of $\vec{d}$.) The work done is the amount of force in the direction of $\vec{d}$, $\|\text{proj}_{\vec{d}}\vec{F}\|$, times $\|d\|$:

$$\begin{aligned}\|\text{proj}_{\vec{d}}\vec{F}\|\cdot\|d\| &= \left\|\frac{\vec{F}\cdot\vec{d}}{\vec{d}\cdot\vec{d}}\vec{d}\right\|\cdot\|d\| \\ &= \left|\frac{\vec{F}\cdot\vec{d}}{\|d\|^2}\right|\cdot\|d\|\cdot\|d\| \\ &= \frac{\left|\vec{F}\cdot\vec{d}\right|}{\|d\|^2}\|d\|^2 \\ &= \left|\vec{F}\cdot\vec{d}\right|.\end{aligned}$$

The expression $\vec{F}\cdot\vec{d}$ will be positive if the angle between $\vec{F}$ and $\vec{d}$ is acute; when the angle is obtuse (hence $\vec{F}\cdot\vec{d}$ is negative), the force is causing motion in the opposite direction of $\vec{d}$, resulting in "negative work." We want to capture this sign, so we drop the absolute value and find that $W=\vec{F}\cdot\vec{d}$.

Definition 60 WORK

Let $\vec{F}$ be a constant force that moves an object in a straight line from point P to point Q. Let $\vec{d}=\vec{PQ}$. The **work** W done by $\vec{F}$ along $\vec{d}$ is $W=\vec{F}\cdot\vec{d}$.

Example 10.3.8: Computing work

A man slides a box along a ramp that rises 3ft over a distance of 15ft by applying 50lb of force as shown in Figure 10.38. Compute the work done.

Solution

The figure indicates that the force applied makes a 30° angle with the horizontal, so $\vec{F}=50\langle\cos 30^\circ,\sin 30^\circ\rangle\approx\langle 43.3,25\rangle$. The ramp is represented by $\vec{d}=\langle 15,3\rangle$. The work done is simply

$$\vec{F}\cdot\vec{d}=50\langle\cos 30^\circ,\sin 30^\circ\rangle\cdot\langle 15,3\rangle\approx 724.5\text{ft--lb}. \tag{10.3.21}$$

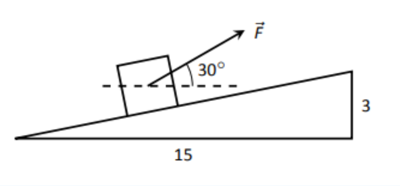


Figure 10.38: Computing work when sliding a box up a ramp in Example 10.3.8.

Note how we did not actually compute the distance the object traveled, nor the magnitude of the force in the direction of travel; this is all inherently computed by the dot product!

The dot product is a powerful way of evaluating computations that depend on angles without actually using angles. The next section explores another "product" on vectors, *the cross product*. Once again, angles play an important role, though in a much different way.

This page titled 10.3: The Dot Product is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

10.4: The Cross Product

"Orthogonality" is immensely important. A quick scan of your current environment will undoubtedly reveal numerous surfaces and edges that are perpendicular to each other (including the edges of this page). The dot product provides a quick test for orthogonality: vectors $\vec{u}$ and $\vec{v}$ are perpendicular if, and only if, $\vec{u}\cdot\vec{v}=0$.

Given two non-parallel, nonzero vectors $\vec{u}$ and $\vec{v}$ in space, it is very useful to find a vector $\vec{w}$ that is perpendicular to both $\vec{u}$ and $\vec{v}$. There is a operation, called the **cross product**, that creates such a vector. This section defines the cross product, then explores its properties and applications.

Definition 61 Cross product

Let $\vec{u}=\langle u_1,u_2,u_3\rangle$ and $\vec{v}=\langle v_1,v_2,v_3\rangle$ be vectors in $\mathbb{R}^3$. The **cross product of $\vec{u}$ and $\vec{v}$**, denoted $\vec{u}\times\vec{v}$, is the vector

$$\vec{u}\times\vec{v}=\langle u_2v_3-u_3v_2,-(u_1v_3-u_3v_1),u_1v_2-u_2v_1\rangle. \tag{10.4.1}$$

This definition can be a bit cumbersome to remember. After an example we will give a convenient method for computing the cross product. For now, careful examination of the products and differences given in the definition should reveal a pattern that is not too difficult to remember. (For instance, in the first component only 2 and 3 appear as subscripts; in the second component, only 1 and 3 appear as subscripts. Further study reveals the order in which they appear.)

Let's practice using this definition by computing a cross product.

Example 10.4.1: Computing a cross product

Let $\vec{u}=\langle 2,-1,4\rangle$ and $\vec{v}=\langle 3,2,5\rangle$. Find $\vec{u}\times\vec{v}$, and verify that it is orthogonal to both $\vec{u}$ and $\vec{v}$.

Solution
Using Definition 61, we have

$$\vec{u}\times\vec{v}=\langle(-1)5-(4)2,-\left((2)5-(4)3\right),(2)2-(-1)3\rangle=\langle -13,2,7\rangle. \tag{10.4.2}$$

(We encourage the reader to compute this product on their own, then verify their result.)

We test whether or not $\vec{u}\times\vec{v}$ is orthogonal to $\vec{u}$ and $\vec{v}$ using the dot product:

$$(\vec{u}\times\vec{v})\cdot\vec{u}=\langle -13,2,7\rangle\cdot\langle 2,-1,4\rangle=0, \tag{10.4.3}$$

$$(\vec{u}\times\vec{v})\cdot\vec{v}=\langle -13,2,7\rangle\cdot\langle 3,2,5\rangle=0. \tag{10.4.4}$$

Since both dot products are zero, $\vec{u}\times\vec{v}$ is indeed orthogonal to both $\vec{u}$ and $\vec{v}$.

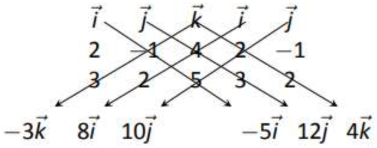
A convenient method of computing the cross product starts with forming a particular 3×3 *matrix*, or rectangular array. The first row comprises the standard unit vectors $\vec{i}$, $\vec{j}$, and $\vec{k}$. The second and third rows are the vectors $\vec{u}$ and $\vec{v}$, respectively. Using $\vec{u}$ and $\vec{v}$ from Example 10.4.1, we begin with:

$$\begin{array}{ccc} \vec{i} & \vec{j} & \vec{k} \\ 2 & -1 & 4 \\ 3 & 2 & 5 \end{array}$$

Now repeat the first two columns after the original three:

$$\begin{array}{ccccc} \vec{i} & \vec{j} & \vec{k} & \vec{i} & \vec{j} \\ 2 & -1 & 4 & 2 & -1 \\ 3 & 2 & 5 & 3 & 2 \end{array}$$

This gives three full "upper left to lower right" diagonals, and three full "upper right to lower left" diagonals, as shown. Compute the products along each diagonal, then add the products on the right and subtract the products on the left:


$\vec{i}$ $\vec{j}$ $\vec{k}$ $\vec{i}$ $\vec{j}$
2 −1 4 2 −1
3 2 5 3 2
$-3\vec{k}$ $8\vec{i}$ $10\vec{j}$
$-5\vec{i}$ $12\vec{j}$ $4\vec{k}$

$$\vec{u}\times\vec{v}=\left(-5\vec{i}+12\vec{j}+4\vec{k}\right)-\left(-3\vec{k}+8\vec{i}+10\vec{j}\right)=-13\vec{i}+2\vec{j}+7\vec{k}=\langle -13,2,7\rangle. \tag{10.4.5}$$

We practice using this method.

Example 10.4.2: Computing a cross product

Let $\vec{u}=\langle 1,3,6\rangle$ and $\vec{v}=\langle -1,2,1\rangle$. Compute both $\vec{u}\times\vec{v}$ and $\vec{v}\times\vec{u}$.

Solution

To compute $\vec{u}\times\vec{v}$, we form the matrix as prescribed above, complete with repeated first columns:

$$\begin{array}{ccccc} \vec{i} & \vec{j} & \vec{k} & \vec{i} & \vec{j} \\ 1 & 3 & 6 & 1 & 3 \\ -1 & 2 & 1 & -1 & 2 \end{array} \tag{10.4.6}$$

We let the reader compute the products of the diagonals; we give the result:

$$\vec{u}\times\vec{v}=\left(3\vec{i}-6\vec{j}+2\vec{k}\right)-\left(-3\vec{k}+12\vec{i}+\vec{j}\right)=\langle -9,-7,5\rangle. \tag{10.4.7}$$

To compute $\vec{v}\times\vec{u}$, we switch the second and third rows of the above matrix, then multiply along diagonals and subtract:

$$\begin{array}{ccccc} \vec{i} & \vec{j} & \vec{k} & \vec{i} & \vec{j} \\ -1 & 2 & 1 & -1 & 2 \\ 1 & 3 & 6 & 1 & 3 \end{array} \tag{10.4.8}$$

Note how with the rows being switched, the products that once appeared on the right now appear on the left, and vice--versa. Thus the result is:

$$\vec{v}\times\vec{u}=\left(12\vec{i}+\vec{j}-3\vec{k}\right)-\left(2\vec{k}+3\vec{i}-6\vec{j}\right)=\langle 9,7,-5\rangle, \tag{10.4.9}$$

which is the opposite of $\vec{u}\times\vec{v}$. We leave it to the reader to verify that each of these vectors is orthogonal to $\vec{u}$ and $\vec{v}$.

Properties of the Cross Product

It is not coincidence that $\vec{v}\times\vec{u}=-(\vec{u}\times\vec{v})$ in the preceding example; one can show using Definition 61 that this will always be the case. The following theorem states several useful properties of the cross product, each of which can be verified by referring to the definition.

THEOREM 87 PROPERTIES OF THE CROSS PRODUCT

Let $\vec{u}$, $\vec{v}$ and $\vec{w}$ be vectors in $\mathbb{R}^3$ and let c be a scalar. The following identities hold:

1. $\vec{u}\times\vec{v}=-(\vec{v}\times\vec{u})$ Anticommutative Property
2. (a) $(\vec{u}+\vec{v})\times\vec{w}=\vec{u}\times\vec{w}+\vec{v}\times\vec{w}$ Distributive Properties
   (b) $\vec{u}\times(\vec{v}+\vec{w})=\vec{u}\times\vec{v}+\vec{u}\times\vec{w}$
3. $c(\vec{u}\times\vec{v})=(c\vec{u})\times\vec{v}=\vec{u}\times(c\vec{v})$
4. (a) $(\vec{u}\times\vec{v})\cdot\vec{u}=0$ Orthogonality Properties
   (b) $(\vec{u}\times\vec{v})\cdot\vec{v}=0$
5. $\vec{u}\times\vec{u}=\vec{0}$
6. $\vec{u}\times\vec{0}=\vec{0}$
7. $\vec{u}\cdot(\vec{v}\times\vec{w})=(\vec{u}\times\vec{v})\cdot\vec{w}$ Triple Scalar Product

We introduced the cross product as a way to find a vector orthogonal to two given vectors, but we did not give a proof that the construction given in Definition 61 satisfies this property. Theorem 87 asserts this property holds; we leave it as a problem in the Exercise section to verify this.

Property 5 from the theorem is also left to the reader to prove in the Exercise section, but it reveals something more interesting than "the cross product of a vector with itself is $\vec{0}$." Let $\vec{u}$ and $\vec{v}$ be parallel vectors; that is, let there be a scalar c such that $\vec{v}=c\vec{u}$. Consider their cross product:

$$\begin{aligned}\vec{u}\times\vec{v}&=\vec{u}\times(c\vec{u})\\&=c(\vec{u}\times\vec{u}) &&\text{(by Property 3 of Theorem 87)}\\&=\vec{0}. &&\text{(by Property 5 of Theorem 87)}\end{aligned}$$

We have just shown that the cross product of parallel vectors is $\vec{0}$. This hints at something deeper. Theorem 86 related the angle between two vectors and their dot product; there is a similar relationship relating the cross product of two vectors and the angle between them, given by the following theorem.

THEOREM 88 THE CROSS PRODUCT AND ANGLES

Let $\vec{u}$ and $\vec{v}$ be vectors in $\mathbb{R}^3$. Then

$$\|\vec{u}\times\vec{v}\|=\|u\|\,\|v\|\sin\theta, \tag{10.4.10}$$

where θ, $0\leq\theta\leq\pi$, is the angle between $\vec{u}$ and $\vec{v}$.

Note: Definition 58 (through Theorem 86) defines $\vec{u}$ and $\vec{v}$ to be orthogonal if $\vec{u}\cdot\vec{v}=0$. We could use Theorem 88 to define $\vec{u}$ and $\vec{v}$ are parallel if $\vec{u}\times\vec{v}=0$. By such a definition, $\vec{0}$ would be both orthogonal and parallel to every vector. Apparent paradoxes such as this are not uncommon in mathematics and can be very useful.

Note that this theorem makes a statement about the *magnitude* of the cross product. When the angle between $\vec{u}$ and $\vec{v}$ is 0 or π (i.e., the vectors are parallel), the magnitude of the cross product is 0. The only vector with a magnitude of 0 is $\vec{0}$ (see Property 9 of Theorem 84), hence the cross product of parallel vectors is $\vec{0}$.

We demonstrate the truth of this theorem in the following example.

Example 10.4.3: The cross product and angles

Let $\vec{u}=\langle 1,3,6\rangle$ and $\vec{v}=\langle -1,2,1\rangle$ as in Example 10.4.2. Verify Theorem 88 by finding θ, the angle between $\vec{u}$ and $\vec{v}$, and the magnitude of $\vec{u}\times\vec{v}$.

Solution

We use Theorem 86 to find the angle between $\vec{u}$ and $\vec{v}$.

$$\begin{aligned}\theta&=\cos^{-1}\left(\frac{\vec{u}\cdot\vec{v}}{\|u\|\,\|v\|}\right)\\&=\cos^{-1}\left(\frac{11}{\sqrt{46}\sqrt{6}}\right)\\&\approx 0.8471=48.54^\circ.\end{aligned}$$

Our work in Example 10.4.2 showed that $\vec{u}\times\vec{v}=\langle -9,-7,5\rangle$, hence $\|\vec{u}\times\vec{v}\|=\sqrt{155}$. Is $\|\vec{u}\times\vec{v}\|=\|u\|\,\|v\|\sin\theta$? Using numerical approximations, we find:

$$\begin{aligned}\|\vec{u}\times\vec{v}\|&=\sqrt{155} & \|u\|\,\|v\|\sin\theta&=\sqrt{46}\sqrt{6}\sin 0.8471\\&\approx 12.45. &&\approx 12.45.\end{aligned}$$

Numerically, they seem equal. Using a right triangle, one can show that

$$\sin\left(\cos^{-1}\left(\frac{11}{\sqrt{46}\sqrt{6}}\right)\right)=\frac{\sqrt{155}}{\sqrt{46}\sqrt{6}}, \tag{10.4.11}$$

> |which allows us to verify the theorem exactly.

Right Hand Rule

The anticommutative property of the cross product demonstrates that $\vec{u} \times \vec{v}$ and $\vec{v} \times \vec{u}$ differ only by a sign -- these vectors have the same magnitude but point in the opposite direction. When seeking a vector perpendicular to $\vec{u}$ and $\vec{v}$, we essentially have two directions to choose from, one in the direction of $\vec{u} \times \vec{v}$ and one in the direction of $\vec{v} \times \vec{u}$. Does it matter which we choose? How can we tell which one we will get without graphing, etc.?

Another wonderful property of the cross product, as defined, is that it follows the **right hand rule**. Given $\vec{u}$ and $\vec{v}$ in $\mathbb{R}^3$ with the same initial point, point the index finger of your right hand in the direction of $\vec{u}$ and let your middle finger point in the direction of $\vec{v}$ (much as we did when establishing the right hand rule for the 3-dimensional coordinate system). Your thumb will naturally extend in the direction of $\vec{u} \times \vec{v}$. One can "practice" this using Figure 10.39. If you switch, and point the index finder in the direction of $\vec{v}$ and the middle finger in the direction of $\vec{u}$, your thumb will now point in the opposite direction, allowing you to "visualize" the anticommutative property of the cross product.

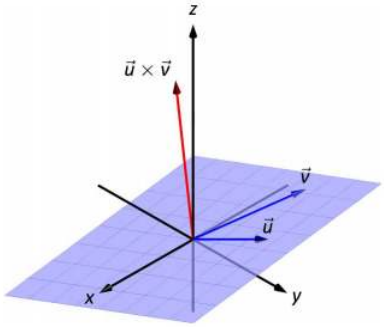


Figure 10.39: Illustrating the Right Hand Rule of the cross product.

Applications of the Cross Product

There are a number of ways in which the cross product is useful in mathematics, physics and other areas of science beyond "just" finding a vector perpendicular to two others. We highlight a few here.

Area of a Parallelogram

It is a standard geometry fact that the area of a parallelogram is $A = bh$, where b is the length of the base and h is the height of the parallelogram, as illustrated in Figure 10.40(a). As shown when defining the Parallelogram Law of vector addition, two vectors $\vec{u}$ and $\vec{v}$ define a parallelogram when drawn from the same initial point, as illustrated in Figure 10.40(b). Trigonometry tells us that $h = \|u\| \sin\theta$, hence the area of the parallelogram is

$$A = \|u\| \, \|v\| \sin\theta = \|\vec{u} \times \vec{v}\|, \tag{10.4.12}$$

where the second equality comes from Theorem 88.

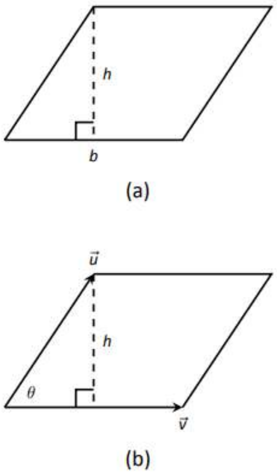


Figure 10.40: Using the cross product to find the area of a parallelogram.

We illustrate using Equation (10.4.12) in the following example.

Example 10.4.4: Finding the area of a parallelogram

1. Find the area of the parallelogram defined by the vectors $\vec{u} = \langle 2, 1\rangle$ and $\vec{v} = \langle 1, 3\rangle$.
2. Verify that the points $A = (1, 1, 1)$, $B = (2, 3, 2)$, $C = (4, 5, 3)$ and $D = (3, 3, 2)$ are the vertices of a parallelogram. Find the area of the parallelogram.

Solution

1. Figure 10.41(a) sketches the parallelogram defined by the vectors $\vec{u}$ and $\vec{v}$. We have a slight problem in that our vectors exist in $\mathbb{R}^2$, not $\mathbb{R}^3$, and the cross product is only defined on vectors in $\mathbb{R}^3$. We skirt this issue by viewing $\vec{u}$ and $\vec{v}$ as vectors in the $x-y$ plane of $\mathbb{R}^3$, and rewrite them as $\vec{u} = \langle 2, 1, 0\rangle$ and $\vec{v} = \langle 1, 3, 0\rangle$. We can now compute the cross product. It is easy to show that $\vec{u} \times \vec{v} = \langle 0, 0, 5\rangle$; therefore the area of the parallelogram is $A = \|\vec{u} \times \vec{v}\| = 5$.
2. To show that the quadrilateral $ABCD$ is a parallelogram (shown in Figure 10.41(b)), we need to show that the opposite sides are parallel. We can quickly show that $\vec{AB} = \vec{DC} = \langle 1, 2, 1\rangle$ and $\vec{BC} = \vec{AD} = \langle 2, 2, 1\rangle$. We find the area by computing the magnitude of the cross product of $\vec{AB}$ and $\vec{BC}$:

$$\vec{AB} \times \vec{BC} = \langle 0, 1, -2\rangle \quad \Rightarrow \quad \|\vec{AB} \times \vec{BC}\| = \sqrt{5} \approx 2.236. \tag{10.4.13}$$

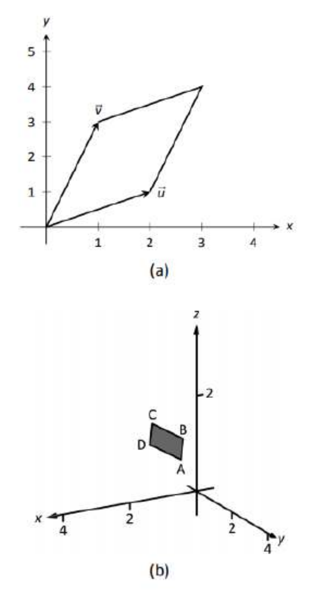


Figure 10.41: Sketching the parallelograms in Example 10.4.4.

This application is perhaps more useful in finding the area of a triangle (in short, triangles are used more often than parallelograms). We illustrate this in the following example.

Example $10.4.5$: Area of a triangle

Find the area of the triangle with vertices $A=(1,2)$, $B=(2,3)$ and $C=(3,1)$, as pictured in Figure 10.42.

Solution

We found the area of this triangle in a previous example to be 1.5 using integration. There we discussed the fact that finding the area of a triangle can be inconvenient using the "$\frac{1}{2}bh$" formula as one has to compute the height, which generally involves finding angles, etc. Using a cross product is much more direct.

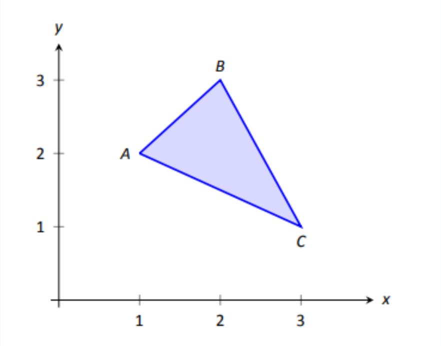


Figure 10.42: Finding the area of a triangle in Example 10.4.5

We can choose any two sides of the triangle to use to form vectors; we choose $\vec{AB}=\langle 1,1\rangle$ and $\vec{AC}=\langle 2,-1\rangle$. As in the previous example, we will rewrite these vectors with a third component of 0 so that we can apply the cross product. The area of the triangle is

$$\frac{1}{2}\|\vec{AB}\times\vec{AC}\|=\frac{1}{2}\|\langle 1,1,0\rangle\times\langle 2,-1,0\rangle\|=\frac{1}{2}\|\langle 0,0,-3\rangle\|=\frac{3}{2}. \tag{10.4.14}$$

We arrive at the same answer as before with less work.

Volume of a Parallelepiped

The three dimensional analogue to the parallelogram is the **parallelepiped**. Each face is parallel to the face opposite face, as illustrated in Figure 10.43. By crossing $\vec{v}$ and $\vec{w}$, one gets a vector whose magnitude is the area of the base. Dotting this vector with $\vec{u}$ computes the volume of parallelepiped! (Up to a sign; take the absolute value.)

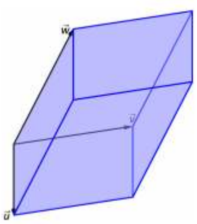


Figure 10.43: A parallelepiped is the three dimensional analogue to the parallelogram.

Thus the volume of a parallelepiped defined by vectors $\vec{u}$, $\vec{v}$ and $\vec{w}$ is

$$V=|\vec{u}\cdot(\vec{v}\times\vec{w})|. \tag{10.4.15}$$

Note how this is the Triple Scalar Product, first seen in Theorem 87. Applying the identities given in the theorem shows that we can apply the Triple Scalar Product in any "order" we choose to find the volume. That is,

$$V=|\vec{u}\cdot(\vec{v}\times\vec{w})|=|\vec{u}\cdot(\vec{w}\times\vec{v})|=|(\vec{u}\times\vec{v})\cdot\vec{w}|, \quad \text{etc.} \tag{10.4.16}$$

Example 10.4.6: Finding the volume of parallelepiped

Find the volume of the parallepiped defined by the vectors $\vec{u} = \langle 1, 1, 0\rangle$, $\vec{v} = \langle -1, 1, 0\rangle$ and $\vec{w} = \langle 0, 1, 1\rangle$.

Solution

We apply Equation (10.4.15). We first find $\vec{v} \times \vec{w} = \langle 1, 1, -1\rangle$. Then

$$|\vec{u} \cdot (\vec{v} \times \vec{w})| = |\langle 1, 1, 0\rangle \cdot \langle 1, 1, -1\rangle| = 2. \tag{10.4.17}$$

So the volume of the parallelepiped is 2 cubic units.

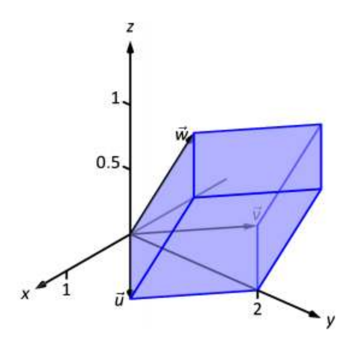


Figure 10.44: A parallelepiped in Example 10.4.6

While this application of the Triple Scalar Product is interesting, it is not used all that often: parallelepipeds are not a common shape in physics and engineering. The last application of the cross product is very applicable in engineering.

Torque

Torque is a measure of the turning force applied to an object. A classic scenario involving torque is the application of a wrench to a bolt. When a force is applied to the wrench, the bolt turns. When we represent the force and wrench with vectors $\vec{F}$ and $\vec{\ell}$, we see that the bolt moves (because of the threads) in a direction orthogonal to $\vec{F}$ and $\vec{\ell}$. Torque is usually represented by the Greek letter τ, or tau, and has units of N·m, a Newton--meter, or ft·lb, a foot--pound.

While a full understanding of torque is beyond the purposes of this book, when a force $\vec{F}$ is applied to a lever arm $\vec{\ell}$, the resulting torque is

$$\vec{\tau} = \vec{\ell} \times \vec{F}. \tag{10.4.18}$$

Example 10.4.7: Computing torque

A lever of length 2ft makes an angle with the horizontal of 45°. Find the resulting torque when a force of 10lb is applied to the end of the level where:

1. the force is perpendicular to the lever, and
2. the force makes an angle of 60° with the lever, as shown in Figure 10.45.

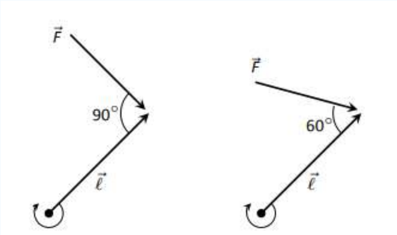


Figure 10.45: Showing a force being applied to a lever in Example 10.4.7.

Solution

1. We start by determining vectors for the force and lever arm. Since the lever arm makes a 45° angle with the horizontal and is 2ft long, we can state that $\vec{\ell} = 2\langle \cos 45^\circ, \sin 45^\circ\rangle = \langle \sqrt{2}, \sqrt{2}\rangle$.

Since the force vector is perpendicular to the lever arm (as seen in the left hand side of Figure 10.45), we can conclude it is making an angle of -45° with the horizontal. As it has a magnitude of 10lb, we can state $\vec{F} = 10\langle \cos(-45^\circ), \sin(-45^\circ)\rangle$ $= \langle 5\sqrt{2}, -5\sqrt{2}\rangle$.

Using Equation 10.45 to find the torque requires a cross product. We again let the third component of each vector be 0 and compute the cross product:

$$\begin{aligned}\vec{\tau} &= \vec{\ell} \times \vec{F} \\ &= \langle \sqrt{2}, \sqrt{2}, 0\rangle \times \langle 5\sqrt{2}, -5\sqrt{2}, 0\rangle \\ &= \langle 0, 0, -20\rangle\end{aligned}$$

This clearly has a magnitude of 20 ft-lb.

We can view the force and lever arm vectors as lying "on the page"; our computation of $\vec{\tau}$ shows that the torque goes "into the page." This follows the Right Hand Rule of the cross product, and it also matches well with the example of the wrench turning the bolt. Turning a bolt clockwise moves it in.

2. Our lever arm can still be represented by $\vec{\ell} = \langle \sqrt{2}, \sqrt{2}\rangle$. As our force vector makes a 60° angle with $\vec{\ell}$, we can see (referencing the right hand side of the figure) that $\vec{F}$ makes a -15° angle with the horizontal. Thus

$$\begin{aligned}\vec{F} = 10\langle \cos -15^\circ, \sin -15^\circ\rangle &= \langle \frac{5(1+\sqrt{3})}{\sqrt{2}}, -\frac{5(1+\sqrt{3})}{\sqrt{2}}\rangle \\ &\approx \langle 9.659, -2.588\rangle.\end{aligned}$$

We again make the third component 0 and take the cross product to find the torque:

$$\begin{aligned}\vec{\tau} &= \vec{\ell} \times \vec{F} \\ &= \langle \sqrt{2}, \sqrt{2}, 0\rangle \times \langle \frac{5(1+\sqrt{3})}{\sqrt{2}}, -\frac{5(1+\sqrt{3})}{\sqrt{2}}, 0\rangle \\ &= \langle 0, 0, -10\sqrt{3}\rangle \\ &\approx \langle 0, 0, -17.321\rangle.\end{aligned}$$

As one might expect, when the force and lever arm vectors *are* orthogonal, the magnitude of force is greater than when the vectors *are not* orthogonal.

While the cross product has a variety of applications (as noted in this chapter), its fundamental use is finding a vector perpendicular to two others. Knowing a vector is orthogonal to two others is of incredible importance, as it allows us to find the equations of lines and planes in a variety of contexts. The importance of the cross product, in some sense, relies on the importance of lines and planes, which see widespread use throughout engineering, physics and mathematics. We study lines and planes in the next two sections.

This page titled 10.4: The Cross Product is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

10.5: Lines

To find the equation of a line in the x-y plane, we need two pieces of information: a point and the slope. The slope conveys *direction* information. As vertical lines have an undefined slope, the following statement is more accurate:

> *To define a line, one needs a point on the line and the direction of the line.*

This holds true for lines in space.

Let P be a point in space, let $\vec{p}$ be the vector with initial point at the origin and terminal point at P (i.e., $\vec{p}$ "points" to P), and let $\vec{d}$ be a vector. Consider the points on the line through P in the direction of $\vec{d}$.

Clearly one point on the line is P; we can say that the *vector* $\vec{p}$ lies at this point on the line. To find another point on the line, we can start at $\vec{p}$ and move in a direction parallel to $\vec{d}$. For instance, starting at $\vec{p}$ and traveling one length of $\vec{d}$ places one at another point on the line. Consider Figure 10.47 where certain points along the line are indicated.

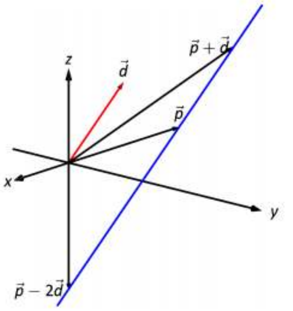


Figure 10.47: Defining a line in space.

The figure illustrates how every point on the line can be obtained by starting with $\vec{p}$ and moving a certain distance in the direction of $\vec{d}$. That is, we can define the line as a function of t:

$$\vec{\ell}(t)=\vec{p}+t\vec{d}. \tag{10.5.1}$$

In many ways, this is *not* a new concept. Compare Equation 10.5.1 to the familiar "$y=mx+b$" equation of a line:

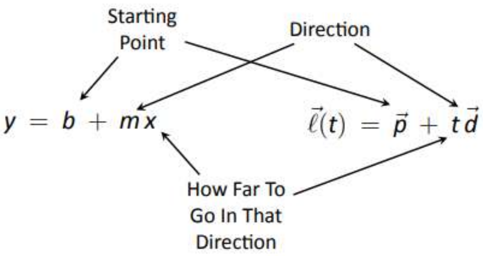

Starting Point
Direction
$y = b + mx$
$\vec{\ell}(t) = \vec{p} + t\vec{d}$
How Far To Go In That Direction

Figure 10.46: Understanding the vector equation of a line.

The equations exhibit the same structure: they give a starting point, define a direction, and state how far in that direction to travel.

Equation 10.5.1 is an example of a **vector-valued function**; the input of the function is a real number and the output is a vector. We will cover vector--valued functions extensively in the next chapter.

There are other ways to represent a line. Let $\vec{p}=\langle x_0,y_0,z_0\rangle$ and let $\vec{d}=\langle a,b,c\rangle$. Then the equation of the line through $\vec{p}$ in the direction of $\vec{d}$ is:

$$\begin{aligned}\vec{\ell}(t)&=\vec{p}+t\vec{d}\\&=\langle x_0,y_0,z_0\rangle+t\langle a,b,c\rangle\\&=\langle x_0+at,y_0+bt,z_0+ct\rangle.\end{aligned}$$

The last line states that the x values of the line are given by $x=x_0+at$, the y values are given by $y=y_0+bt$, and the z values are given by $z=z_0+ct$. These three equations, taken together, are the **parametric equations of the line** through $\vec{p}$ in the direction of $\vec{d}$.

Finally, each of the equations for x, y and z above contain the variable t. We can solve for t in each equation:

$$\begin{aligned} x &= x_0 + at & \Rightarrow \quad t &= \frac{x - x_0}{a}, \\ y &= y_0 + bt & \Rightarrow \quad t &= \frac{y - y_0}{b}, \\ z &= z_0 + ct & \Rightarrow \quad t &= \frac{z - z_0}{c}, \end{aligned}$$

assuming $a, b, c \neq 0$.

Since t is equal to each expression on the right, we can set these equal to each other, forming the **symmetric equations of the line** through $\vec{p}$ in the direction of $\vec{d}$:

$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}. \tag{10.5.2}$$

Each representation has its own advantages, depending on the context. We summarize these three forms in the following definition, then give examples of their use.

Definition 62 Equations of Lines in Space

Consider the line in space that passes through $\vec{p} = \langle x_0, y_0, z_0 \rangle$ in the direction of $\vec{d} = \langle a, b, c \rangle$.

1. The **vector equation** of the line is

$$\vec{\ell}(t) = \vec{p} + t\vec{d}. \tag{10.5.3}$$

2. The **parametric equations** of the line are

$$x = x_0 + at, \quad y = y_0 + bt, \quad z = z_0 + ct. \tag{10.5.4}$$

3. The **symmetric equations** of the line are

$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}. \tag{10.5.5}$$

Example 10.5.1: Finding the equation of a line

Give all three equations, as given in Definition 62, of the line through $P = (2, 3, 1)$ in the direction of $\vec{d} = \langle -1, 1, 2 \rangle$. Does the point $Q = (-1, 6, 6)$ lie on this line?

Solution

We identify the point $P = (2, 3, 1)$ with the vector $\vec{p} = \langle 2, 3, 1 \rangle$. Following the definition, we have

1. the vector equation of the line is $\vec{\ell}(t) = \langle 2, 3, 1 \rangle + t\langle -1, 1, 2 \rangle$;
2. the parametric equations of the line are

$$x = 2 - t, \quad y = 3 + t, \quad z = 1 + 2t; \text{ and} \tag{10.5.6}$$

3. the symmetric equations of the line are

$$\frac{x - 2}{-1} = \frac{y - 3}{1} = \frac{z - 1}{2}. \tag{10.5.7}$$

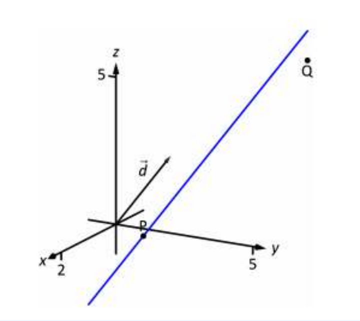


Figure 10.48: Graphing a line in Example 10.5.1.

The first two equations of the line are useful when a t value is given: one can immediately find the corresponding point on the line. These forms are good when calculating with a computer; most software programs easily handle equations in these formats. (For instance, to make Figure 10.48, a certain graphics program was given the input `(2-x,3+x,1+2*x)`. This particular program requires the variable always be "x" instead of "t").

Does the point $Q = (-1, 6, 6)$ lie on the line? The graph in Figure 10.48 makes it clear that it does not. We can answer this question without the graph using any of the three equation forms. Of the three, the symmetric equations are probably best suited for this task. Simply plug in the values of x, y and z and see if equality is maintained:

$$\frac{-1-2}{-1} \overset{?}{=} \frac{6-3}{1} \overset{?}{=} \frac{6-1}{2} \quad \Rightarrow \quad 3 = 3 \neq 2.5. \tag{10.5.8}$$

We see that Q does not lie on the line as it did not satisfy the symmetric equations.

Example $10.5.2$: Finding the equation of a line through two points

Find the parametric equations of the line through the points $P = (2, -1, 2)$ and $Q = (1, 3, -1)$.

Solution
Recall the statement made at the beginning of this section: to find the equation of a line, we need a point and a direction. We have *two* points; either one will suffice. The direction of the line can be found by the vector with initial point P and terminal point Q: $\vec{PQ} = \langle -1, 4, -3 \rangle$.

The parametric equations of the line ℓ through P in the direction of $\vec{PQ}$ are:

$$\ell: \quad x = 2 - t \quad y = -1 + 4t \quad z = 2 - 3t. \tag{10.5.9}$$

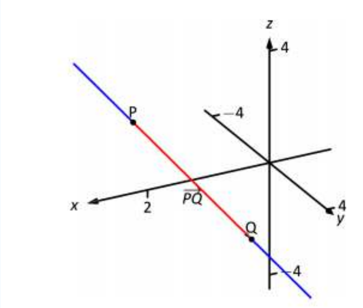
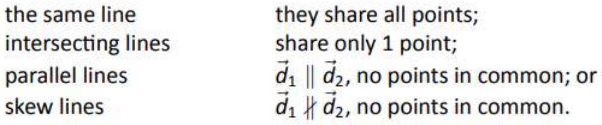


Figure 10.49: A graph of the line in Example 10.5.2.

A graph of the points and line are given in Figure 10.49. Note how in the given parametrization of the line, $t = 0$ corresponds to the point P, and $t = 1$ corresponds to the point Q. This relates to the understanding of the vector equation of a line described in Figure 10.46. The parametric equations "start" at the point P, and t determines how far in the direction of $\vec{PQ}$ to travel. When $t = 0$, we travel 0 lengths of $\vec{PQ}$; when $t = 1$, we travel one length of $\vec{PQ}$, resulting in the point Q.

Parallel, Intersecting and Skew Lines

In the plane, two *distinct* lines can either be parallel or they will intersect at exactly one point. In space, given equations of two lines, it can sometimes be difficult to tell whether the lines are distinct or not (i.e., the same line can be represented in different ways). Given lines $\vec{\ell}_1(t) = \vec{p}_1 + t\vec{d}_1$ and $\vec{\ell}_2(t) = \vec{p}_2 + t\vec{d}_2$, we have four possibilities: $\vec{\ell}_1$ and $\vec{\ell}_2$ are



the same line	they share all points;
intersecting lines	share only 1 point;
parallel lines	$\vec{d}_1 \parallel \vec{d}_2$, no points in common; or
skew lines	$\vec{d}_1 \nparallel \vec{d}_2$, no points in common.

The next two examples investigate these possibilities.

Example $10.5.3$: Comparing lines

Consider lines ℓ_1 and ℓ_2, given in parametric equation form:

$$\ell_1 : \begin{array}{rcl} x &=& 1+3t \\ y &=& 2-t \\ z &=& t \end{array} \qquad \ell_2 : \begin{array}{rcl} x &=& -2+4s \\ y &=& 3+s \\ z &=& 5+2s. \end{array} \tag{10.5.10}$$

Determine whether ℓ_1 and ℓ_2 are the same line, intersect, are parallel, or skew.

Solution

We start by looking at the directions of each line. Line ℓ_1 has the direction given by $\vec{d}_1 = \langle 3, -1, 1\rangle$ and line ℓ_2 has the direction given by $\vec{d}_2 = \langle 4, 1, 2\rangle$. It should be clear that $\vec{d}_1$ and $\vec{d}_2$ are not parallel, hence ℓ_1 and ℓ_2 are not the same line, nor are they parallel. Figure 10.50 verifies this fact (where the points and directions indicated by the equations of each line are identified).

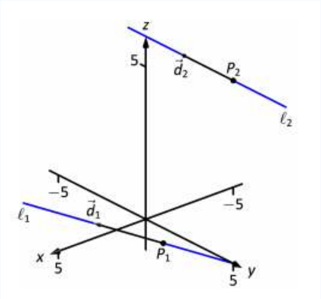


Figure 10.50: Sketching the lines from Example 10.5.3

We next check to see if they intersect (if they do not, they are skew lines). To find if they intersect, we look for t and s values such that the respective x, y and z values are the same. That is, we want s and t such that:

$$\begin{array}{rcl} 1+3t &=& -2+4s \\ 2-t &=& 3+s \\ t &=& 5+2s. \end{array} \tag{10.5.11}$$

This is a relatively simple system of linear equations. Since the last equation is already solved for t, substitute that value of t into the equation above it:

$$2-(5+2s) = 3+s \quad \Rightarrow \quad s=-2,\ t=1. \tag{10.5.12}$$

A key to remember is that we have *three* equations; we need to check if $s=-2,\ t=1$ satisfies the first equation as well:

$$1+3(1) \neq -2+4(-2). \tag{10.5.13}$$

It does not. Therefore, we conclude that the lines ℓ_1 and ℓ_2 are skew.

Example $10.5.4$: Comparing lines

Consider lines ℓ_1 and ℓ_2, given in parametric equation form:

$$\ell_1 : \begin{array}{rcl} x &=& -0.7+1.6t \\ y &=& 4.2+2.72t \\ z &=& 2.3-3.36t \end{array} \qquad \ell_2 : \begin{array}{rcl} x &=& 2.8-2.9s \\ y &=& 10.15-4.93s \\ z &=& -5.05+6.09s. \end{array} \tag{10.5.14}$$

Determine whether ℓ_1 and ℓ_2 are the same line, intersect, are parallel, or skew.

Solution

It is obviously very difficult to simply look at these equations and discern anything. This is done intentionally. In the "real world," most equations that are used do not have nice, integer coefficients. Rather, there are lots of digits after the decimal and the equations can look "messy."

We again start by deciding whether or not each line has the same direction. The direction of ℓ_1 is given by $\vec{d}_1 = \langle 1.6, 2.72, -3.36\rangle$ and the direction of ℓ_2 is given by $\vec{d}_2 = \langle -2.9, -4.93, 6.09\rangle$. When it is not clear through observation whether two vectors are parallel or not, the standard way of determining this is by comparing their respective unit vectors. Using a calculator, we find:

$$\begin{aligned}\vec{u}_1 &= \frac{\vec{d}_1}{\|\vec{d}_1\|} = \langle 0.3471, 0.5901, -0.7289\rangle \\ \vec{u}_2 &= \frac{\vec{d}_2}{\|\vec{d}_2\|} = \langle -0.3471, -0.5901, 0.7289\rangle.\end{aligned}$$

The two vectors seem to be parallel (at least, their components are equal to 4 decimal places). In most situations, it would suffice to conclude that the lines are at least parallel, if not the same. One way to be sure is to rewrite $\vec{d}_1$ and $\vec{d}_2$ in terms of fractions, not decimals. We have

$$\vec{d}_1 = \langle \frac{16}{10}, \frac{272}{100}, -\frac{336}{100}\rangle \qquad \vec{d}_2 = \langle -\frac{29}{10}, -\frac{493}{100}, \frac{609}{100}\rangle. \tag{10.5.15}$$

One can then find the magnitudes of each vector in terms of fractions, then compute the unit vectors likewise. After a lot of manual arithmetic (or after briefly using a computer algebra system), one finds that

$$\vec{u}_1 = \langle \sqrt{\frac{10}{83}}, \frac{17}{\sqrt{830}}, -\frac{21}{\sqrt{830}}\rangle \qquad \vec{u}_2 = \langle -\sqrt{\frac{10}{83}}, -\frac{17}{\sqrt{830}}, \frac{21}{\sqrt{830}}\rangle. \tag{10.5.16}$$

We can now say without equivocation that these lines are parallel.

Are they the same line? The parametric equations for a line describe one point that lies on the line, so we know that the point $P_1 = (-0.7, 4.2, 2.3)$ lies on ℓ_1. To determine if this point also lies on ℓ_2, plug in the x, y and z values of P_1 into the symmetric equations for ℓ_2:

$$\begin{aligned}\frac{(-0.7)-2.8}{-2.9} \overset{?}{=} \frac{(4.2)-10.15}{-4.93} &\overset{?}{=} \frac{(2.3)-(-5.05)}{6.09} \\ \Rightarrow \quad 1.2069 = 1.2069 &= 1.2069.\end{aligned} \tag{10.5.17}$$

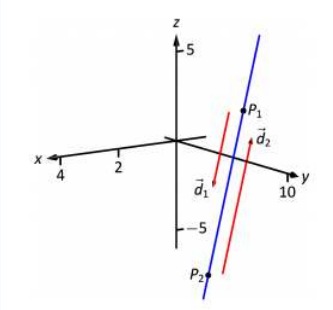


Figure 10.51: Graphing the lines in Example 10.5.4

The point P_1 lies on both lines, so we conclude they are the same line, just parametrized differently. Figure 10.51 graphs this line along with the points and vectors described by the parametric equations. Note how $\vec{d}_1$ and $\vec{d}_2$ are parallel, though point in opposite directions (as indicated by their unit vectors above).

Distances

Given a point Q and a line $\vec{\ell}(t) = \vec{p} + t\vec{d}$ in space, it is often useful to know the distance from the point to the line. (Here we use the standard definition of "distance," i.e., the length of the shortest line segment from the point to the line.) Identifying $\vec{p}$ with the point P, Figure 10.52 will help establish a general method of computing this distance h.

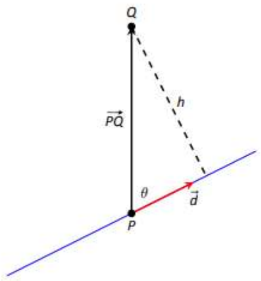


Figure 10.52: Establishing the distance from a point to a line.

From trigonometry, we know $h = \|\vec{PQ}\| \sin\theta$. We have a similar identity involving the cross product: $\|\vec{PQ} \times \vec{d}\| = \|\vec{PQ}\|\,\|d\|\sin\theta$. Divide both sides of this latter equation by $\|d\|$ to obtain h:

$$h = \frac{\|\vec{PQ} \times \vec{d}\|}{\|d\|}. \tag{10.5.18}$$

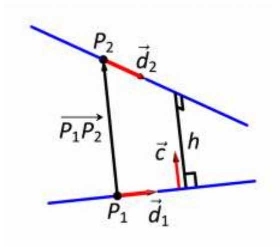


Figure 10.53: Establishing the distance between lines.

It is also useful to determine the distance between lines, which we define as the length of the shortest line segment that connects the two lines (an argument from geometry shows that this line segments is perpendicular to both lines). Let lines $\vec{\ell}_1(t) = \vec{p}_1 + t\vec{d}_1$ and $\vec{\ell}_2(t) = \vec{p}_2 + t\vec{d}_2$ be given, as shown in Figure 10.53. To find the direction orthogonal to both $\vec{d}_1$ and $\vec{d}_2$, we take the cross product: $\vec{c} = \vec{d}_1 \times \vec{d}_2$. The magnitude of the orthogonal projection of $\vec{P_1P_2}$ onto $\vec{c}$ is the distance h we seek:

$$\begin{aligned} h &= \|\text{proj}_{\vec{c}}\, \vec{P_1P_2}\| \\ &= \|\frac{\vec{P_1P_2} \cdot \vec{c}}{\vec{c} \cdot \vec{c}}\vec{c}\| \\ &= \frac{|\vec{P_1P_2} \cdot \vec{c}|}{\|c\|^2}\|c\| \\ &= \frac{|\vec{P_1P_2} \cdot \vec{c}|}{\|c\|}. \end{aligned}$$

A problem in the Exercise section is to show that this distance is 0 when the lines intersect. Note the use of the Triple Scalar Product: $\vec{P_1P_2} \cdot c = \vec{P_1P_2} \cdot (\vec{d}_1 \times \vec{d}_2)$.

The following Key Idea restates these two distance formulas.

KEY IDEA 50 DISTANCE TO LINES

1. Let P be a point on a line ℓ that is parallel to $\vec{d}$. The distance h from a point Q to the line ℓ is:

$$h = \frac{\|\vec{PQ} \times \vec{d}\|}{\|d\|}. \tag{10.5.19}$$

2. Let P_1 be a point on line ℓ_1 that is parallel to $\vec{d}_1$, and let P_2 be a point on line ℓ_2 parallel to $\vec{d}_2$, and let $\vec{c} = \vec{d}_1 \times \vec{d}_2$, where lines ℓ_1 and ℓ_2 are not parallel. The distance h between the two lines is:

$$h = \frac{|\vec{P_1P_2} \cdot \vec{c}|}{\|c\|}. \tag{10.5.20}$$

Example 10.5.5: Finding the distance from a point to a line

Find the distance from the point $Q = (1, 1, 3)$ to the line $\vec{\ell}\,(t) = \langle 1, -1, 1\rangle + t\langle 2, 3, 1\rangle$.

Solution

The equation of the line gives us the point $P = (1, -1, 1)$ that lies on the line, hence $\vec{PQ} = \langle 0, 2, 2\rangle$. The equation also gives $\vec{d} = \langle 2, 3, 1\rangle$. Following Key Idea 50, we have the distance as

$$\begin{aligned} h &= \frac{\|\vec{PQ} \times \vec{d}\,\|}{\|d\|} \\ &= \frac{\|\langle -4, 4, -4\rangle\|}{\sqrt{14}} \\ &= \frac{4\sqrt{3}}{\sqrt{14}} \approx 1.852. \end{aligned}$$

The point Q is approximately 1.852 units from the line $\vec{\ell}\,(t)$.

Example 10.5.6: Finding the distance between lines

Find the distance between the lines

$$\ell_1 : \begin{array}{rcl} x &=& 1+3t \\ y &=& 2-t \\ z &=& t \end{array} \qquad \ell_2 : \begin{array}{rcl} x &=& -2+4s \\ y &=& 3+s \\ z &=& 5+2s. \end{array} \tag{10.5.21}$$

Solution

These are the sames lines as given in Example 10.5.3, where we showed them to be skew. The equations allow us to identify the following points and vectors:

$$P_1 = (1, 2, 0) \quad P_2 = (-2, 3, 5) \quad \Rightarrow \quad \vec{P_1P_2} = \langle -3, 1, 5\rangle. \tag{10.5.22}$$

$$\vec{d}_1 = \langle 3, -1, 1\rangle \quad \vec{d}_2 = \langle 4, 1, 2\rangle \quad \Rightarrow \quad \vec{c} = \vec{d}_1 \times \vec{d}_2 = \langle -3, -2, 7\rangle. \tag{10.5.23}$$

From Key Idea 50 we have the distance h between the two lines is

$$\begin{aligned} h &= \frac{|\vec{P_1P_2} \cdot \vec{c}\,|}{\|c\|} \\ &= \frac{42}{\sqrt{62}} \approx 5.334. \end{aligned}$$

The lines are approximately 5.334 units apart.

One of the key points to understand from this section is this: to describe a line, we need a point and a direction. Whenever a problem is posed concerning a line, one needs to take whatever information is offered and glean point and direction information. Many questions can be asked (and *are* asked in the Exercise section) whose answer immediately follows from this understanding.

Lines are one of two fundamental objects of study in space. The other fundamental object is the *plane*, which we study in detail in the next section. Many complex three dimensional objects are studied by approximating their surfaces with lines and planes.

This page titled 10.5: Lines is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

10.6: Planes

Any flat surface, such as a wall, table top or stiff piece of cardboard can be thought of as representing part of a plane. Consider a piece of cardboard with a point P marked on it. One can take a nail and stick it into the cardboard at P such that the nail is perpendicular to the cardboard; see Figure 10.54.

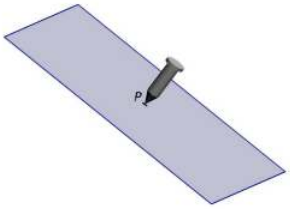


Figure 10.54: Illustrating defining a plane with a sheet of cardboard and a nail.

This nail provides a "handle" for the cardboard. Moving the cardboard around moves P to different locations in space. Tilting the nail (but keeping P fixed) tilts the cardboard. Both moving and tilting the cardboard defines a different plane in space. In fact, we can define a plane by: 1) the location of P in space, and 2) the direction of the nail.

The previous section showed that one can define a line given a point on the line and the direction of the line (usually given by a vector). One can make a similar statement about planes: we can define a plane in space given a point on the plane and the direction the plane "faces" (using the description above, the direction of the nail). Once again, the direction information will be supplied by a vector, called a **normal vector**, that is orthogonal to the plane.

What exactly does "orthogonal to the plane" mean? Choose any two points P and Q in the plane, and consider the vector $\vec{PQ}$. We say a vector $\vec{n}$ is orthogonal to the plane if $\vec{n}$ is perpendicular to $\vec{PQ}$ for all choices of P and Q; that is, if $\vec{n}\cdot\vec{PQ}=0$ for all P and Q.

This gives us way of writing an equation describing the plane. Let $P=(x_0,y_0,z_0)$ be a point in the plane and let $\vec{n}=\langle a,b,c\rangle$ be a normal vector to the plane. A point $Q=(x,y,z)$ lies in the plane defined by P and $\vec{n}$ if, and only if, $\vec{PQ}$ is orthogonal to $\vec{n}$. Knowing $\vec{PQ}=\langle x-x_0,y-y_0,z-z_0\rangle$, consider:

$$\begin{aligned}\vec{PQ}\cdot\vec{n}&=0\\ \langle x-x_0,y-y_0,z-z_0\rangle\cdot\langle a,b,c\rangle&=0\\ a(x-x_0)+b(y-y_0)+c(z-z_0)&=0\end{aligned}\tag{10.6.1}$$

Equation 10.6.1 defines an *implicit* function describing the plane. More algebra produces:

$$ax+by+cz=ax_0+by_0+cz_0.$$

The right hand side is just a number, so we replace it with d:

$$ax+by+cz=d.\tag{10.6.2}$$

As long as $c\neq 0$, we can solve for z:

$$z=\frac{1}{c}(d-ax-by).\tag{10.6.3}$$

Equation 10.6.3 is especially useful as many computer programs can graph functions in this form. Equations 10.6.1 and 10.6.2 have specific names, given next.

Definition 63 Equations of a Plane in Standard and General Forms

$$a(x-x_0)+b(y-y_0)+c(z-z_0)=0;\tag{10.6.4}$$

the equation's **general form** is

$$ax+by+cz=d.\tag{10.6.5}$$

A key to remember throughout this section is this: to find the equation of a plane, we need a point and a normal vector. We will give several examples of finding the equation of a plane, and in each one different types of information are given. In each case, we need to use the given information to find a point on the plane and a normal vector.

Example 10.6.1: Finding the equation of a plane.

Write the equation of the plane that passes through the points $P=(1,1,0)$, $Q=(1,2,-1)$ and $R=(0,1,2)$ in standard form.

Solution
We need a vector $\vec{n}$ that is orthogonal to the plane. Since P, Q and R are in the plane, so are the vectors $\vec{PQ}$ and $\vec{PR}$; $\vec{PQ}\times\vec{PR}$ is orthogonal to $\vec{PQ}$ and $\vec{PR}$ and hence the plane itself.

It is straightforward to compute $\vec{n}=\vec{PQ}\times\vec{PR}=\langle 2,1,1\rangle$. We can use any point we wish in the plane (any of P, Q or R will do) and we arbitrarily choose P. Following Definition 63, the equation of the plane in standard form is

$$2(x-1)+(y-1)+z=0. \tag{10.6.6}$$

The plane is sketched in Figure 10.55.

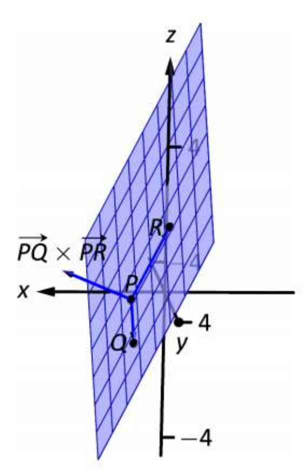


Figure 10.55: Sketching the plane in Example 10.6.1

We have just demonstrated the fact that any three non-collinear points define a plane. (This is why a three-legged stool does not "rock;" it's three feet always lie in a plane. A four-legged stool will rock unless all four feet lie in the same plane.)\

Example 10.6.2: Finding the equation of a plane.

Verify that lines ℓ_1 and ℓ_2, whose parametric equations are given below, intersect, then give the equation of the plane that contains these two lines in general form.

$$\ell_1: \begin{array}{rcl} x &=& -5+2s \\ y &=& 1+s \\ z &=& -4+2s \end{array} \qquad \ell_2: \begin{array}{rcl} x &=& 2+3t \\ y &=& 1-2t \\ z &=& 1+t \end{array} \tag{10.6.7}$$

Solution
The lines clearly are not parallel. If they do not intersect, they are skew, meaning there is not a plane that contains them both. If they do intersect, there is such a plane.

To find their point of intersection, we set the x, y and z equations equal to each other and solve for s and t:

$$\begin{array}{rcl} -5+2s &=& 2+3t \\ 1+s &=& 1-2t \\ -4+2s &=& 1+t \end{array} \quad \Rightarrow \quad s=2, \quad t=-1. \tag{10.6.8}$$

When $s=2$ and $t=-1$, the lines intersect at the point $P=(-1,3,0)$.

Let $\vec{d}_1 = \langle 2,1,2\rangle$ and $\vec{d}_2 = \langle 3,-2,1\rangle$ be the directions of lines ℓ_1 and ℓ_2, respectively. A normal vector to the plane containing these the two lines will also be orthogonal to $\vec{d}_1$ and $\vec{d}_2$. Thus we find a normal vector $\vec{n}$ by computing $\vec{n} = \vec{d}_1 \times \vec{d}_2 = \langle 5, 4-7\rangle$.

We can pick any point in the plane with which to write our equation; each line gives us infinite choices of points. We choose P, the point of intersection. We follow Definition 63 to write the plane's equation in general form:

$$\begin{aligned} 5(x+1)+4(y-3)-7z &= 0 \\ 5x+5+4y-12-7z &= 0 \\ 5x+4y-7z &= 7. \end{aligned}$$

The plane's equation in general form is $5x+4y-7z=7$; it is sketched in Figure 10.56.

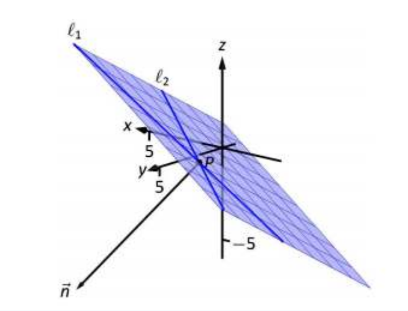


Figure 10.56: Sketching the plane in Example 10.6.2

Example 10.6.3: Finding the equation of a plane

Give the equation, in standard form, of the plane that passes through the point $P = (-1,0,1)$ and is orthogonal to the line with vector equation $\vec{\ell}(t) = \langle -1,0,1\rangle + t\langle 1,2,2\rangle$.

Solution

As the plane is to be orthogonal to the line, the plane must be orthogonal to the direction of the line given by $\vec{d} = \langle 1,2,2\rangle$. We use this as our normal vector. Thus the plane's equation, in standard form, is

$$(x+1)+2y+2(z-1)=0. \tag{10.6.9}$$

The line and plane are sketched in Figure 10.57.

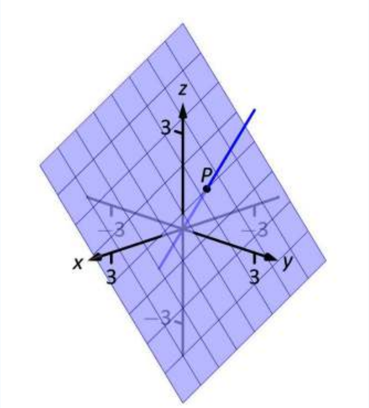


Figure 10.57: The line and plane in Example 10.6.3

Example 10.6.4: Finding the intersection of two planes

Give the parametric equations of the line that is the intersection of the planes p_1 and p_2, where:

$$\begin{aligned} p_1 &: x-(y-2)+(z-1)=0 \\ p_2 &: -2(x-2)+(y+1)+(z-3)=0 \end{aligned}$$

Solution

To find an equation of a line, we need a point on the line and the direction of the line.

We can find a point on the line by solving each equation of the planes for z:

$$\begin{gathered} p_1 : z = -x + y - 1 \\ p_2 : z = 2x - y - 2 \end{gathered}$$

We can now set these two equations equal to each other (i.e., we are finding values of x and y where the planes have the same z value):

$$\begin{aligned} -x + y - 1 &= 2x - y - 2 \\ 2y &= 3x - 1 \\ y &= \frac{1}{2}(3x - 1) \end{aligned}$$

We can choose any value for x; we choose $x = 1$. This determines that $y = 1$. We can now use the equations of either plane to find z: when $x = 1$ and $y = 1$, $z = -1$ on both planes. We have found a point P on the line: $P = (1, 1, -1)$.

We now need the direction of the line. Since the line lies in each plane, its direction is orthogonal to a normal vector for each plane. Considering the equations for p_1 and p_2, we can quickly determine their normal vectors. For p_1, $\vec{n}_1 = \langle 1, -1, 1\rangle$ and for p_2, $\vec{n}_2 = \langle -2, 1, 1\rangle$. A direction orthogonal to both of these directions is their cross product: $\vec{d} = \vec{n}_1 \times \vec{n}_2 = \langle -2, -3, -1\rangle$.

The parametric equations of the line through $P = (1, 1, -1)$ in the direction of $d = \langle -2, -3, -1\rangle$ is:

$$\ell : \quad x = -2t + 1 \quad y = -3t + 1 \quad z = -t - 1. \tag{10.6.10}$$

The planes and line are graphed in Figure 10:58.

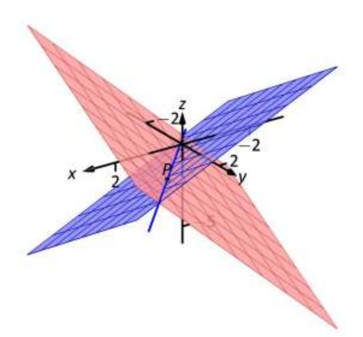


Figure 10.58: Graphing the planes and their line of intersection in Example 10.6.4

Example 10.6.5: Finding the intersection of a plane and a line

Find the point of intersection, if any, of the line $\ell(t) = \langle 3, -3, -1\rangle + t\langle -1, 2, 1\rangle$ and the plane with equation in general form $2x + y + z = 4$.

Solution
The equation of the plane shows that the vector $\vec{n} = \langle 2, 1, 1\rangle$ is a normal vector to the plane, and the equation of the line shows that the line moves parallel to $\vec{d} = \langle -1, 2, 1\rangle$. Since these are not orthogonal, we know there is a point of intersection. (If there were orthogonal, it would mean that the plane and line were parallel to each other, either never intersecting or the line was in the plane itself.)

To find the point of intersection, we need to find a t value such that $\ell(t)$ satisfies the equation of the plane. Rewriting the equation of the line with parametric equations will help:

$$\ell(t) = \begin{cases} x = 3 - t \\ y = -3 + 2t \\ z = -1 + t \end{cases}. \tag{10.6.11}$$

Replacing x, y and z in the equation of the plane with the expressions containing t found in the equation of the line allows us to determine a t value that indicates the point of intersection:

$$\begin{aligned} 2x + y + z &= 4 \\ 2(3 - t) + (-3 + 2t) + (-1 + t) &= 4 \\ t &= 2. \end{aligned}$$

When $t = 2$, the point on the line satisfies the equation of the plane; that point is $\ell(2) = \langle 1, 1, 1\rangle$. Thus the point $(1, 1, 1)$ is the point of intersection between the plane and the line, illustrated in Figure 10.59.

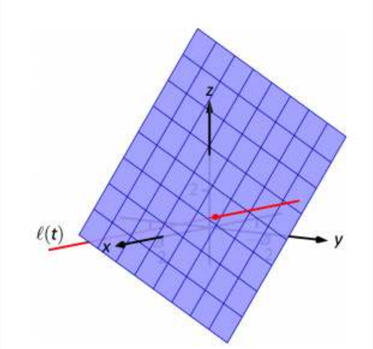


Figure 10.59: Illustrating the intersection of a line and a plane in Example 10.6.5

Distances

Just as it was useful to find distances between points and lines in the previous section, it is also often necessary to find the distance from a point to a plane.

Consider Figure 10.60, where a plane with normal vector $\vec{n}$ is sketched containing a point P and a point Q, not on the plane, is given. We measure the distance from Q to the plane by measuring the length of the projection of $\vec{PQ}$ onto $\vec{n}$. That is, we want:

$$\|\text{proj}_{\,\vec{n}}\,\vec{PQ}\| = \|\frac{\vec{n}\cdot\vec{PQ}}{\|n\|^2}\vec{n}\| = \frac{|\vec{n}\cdot\vec{PQ}|}{\|n\|} \tag{10.6.12}$$

Equation 10.6.12 is important as it does more than just give the distance between a point and a plane. We will see how it allows us to find several other distances as well: the distance between parallel planes and the distance from a line and a plane. Because Equation 10.6.12 is important, we restate it as a Key Idea.

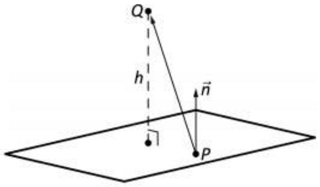


Figure 10.60: *Illustrating finding the distance from a point to a plane.*

KEY IDEA 51 Distance from a Point to a Plane

Let a plane with normal vector $\vec{n}$ be given, and let Q be a point. The distance h from Q to the plane is

$$h = \frac{|\vec{n}\cdot\vec{PQ}|}{\|n\|}, \tag{10.6.13}$$

where P is any point in the plane.

Example 10.6.6: Distance between a point and a plane

Find the distance bewteen the point $Q = (2, 1, 4)$ and the plane with equation $2x - 5y + 6z = 9$.

Solution

Using the equation of the plane, we find the normal vector $\vec{n} = \langle 2, -5, 6\rangle$. To find a point on the plane, we can let x and y be anything we choose, then let z be whatever satisfies the equation. Letting x and y be 0 seems simple; this makes $z = 1.5$. Thus we let $P = \langle 0, 0, 1.5\rangle$, and $\vec{PQ} = \langle 2, 1, 2.5\rangle$.

The distance h from Q to the plane is given by Key Idea 51:

$$\begin{aligned} h &= \frac{|\vec{n}\cdot\vec{PQ}|}{\|n\|} \\ &= \frac{|\langle 2,-5,6\rangle\cdot\langle 2,1,2.5\rangle|}{\|\langle 2,-5,6\rangle\|} \\ &= \frac{|14|}{\sqrt{65}} \\ &\approx 1.74. \end{aligned}$$

We can use Key Idea 51 to find other distances. Given two parallel planes, we can find the distance between these planes by letting P be a point on one plane and Q a point on the other. If ℓ is a line parallel to a plane, we can use the Key Idea to find the distance between them as well: again, let P be a point in the plane and let Q be any point on the line. (One can also use Key Idea 50.) The Exercise section contains problems of these types.

These past two sections have not explored lines and planes in space as an exercise of mathematical curiosity. However, there are many, many applications of these fundamental concepts. Complex shapes can be modeled (or, *approximated*) using planes. For instance, part of the exterior of an aircraft may have a complex, yet smooth, shape, and engineers will want to know how air flows across this piece as well as how heat might build up due to air friction. Many equations that help determine air flow and heat dissipation are difficult to apply to arbitrary surfaces, but simple to apply to planes. By approximating a surface with millions of small planes one can more readily model the needed behavior.

This page titled 10.6: Planes is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

10.E: Applications of Vectors (Exercises)

10.1: Introduction to Cartesian Coordinates in Space

Terms and Concepts

1. Axes drawn in space must conform to the ________ _________ rule.

2. In the plane, the equation $x = 2$ defines a ________; in space, $x = 2$ defines a ________.

3. In the plane, the equation $y = x^2$ defines a ________; in space, $y = x^2$ defines a ________.

4. Which quadric surface looks like a Pringles chip?

5. Consider the hyperbola $x^2 - y^2 = 1$ in the plane. If this hyperbola is rotated about the x-axis, what quadric surface is formed?

6. Consider the hyperbola $x^2 - y^2 = 1$ in the plane. If this hyperbola is rotated about the y-axis, what quadric surface is formed?

Problems

7. The points $A = (1, 4, 2)$, $B = (2, 6, 3)$ and $C = (4, 3, 1)$ form a triangle in space. Find the distances between each pair of points and determine if the triangle is a right triangle.

8. The points $A = (1, 1, 3)$, $B = (3, 2, 7)$, $C = (2, 0, 8)$ and $D = (0, -1, 4)$ form a quadrilateral ABCD in space. Is this a parallelogram?

9. Find the center and radius of the sphere defined by $x^2 - 8x + y^2 + 2y + z^2 + 8 = 0$.

10. Find the center and radius of the sphere defined by $x^2 + y^2 + z^2 + 4x - 2y - 4z + 4 = 0$.

In Exercises 11-14, describe the region in space defined by the inequalities.

11. $x^2 + y^2 + z^2 < 1$

12. $0 \le x \le 3$

13. $x \ge 0, y \ge 0, z \ge 0$

14. $y \ge 3$

In Exercises 15-18, sketch the cylinder in space.

15. $z = x^3$

16. $y = \cos z$

17. $\frac{x^2}{4} + \frac{y^2}{9} = 1$

18. $y = \frac{1}{x}$

In Exercises 19-22, give the equation of the surface of revolution described.

19. Revolve $z = \frac{1}{1+y^2}$ about the y-axis.

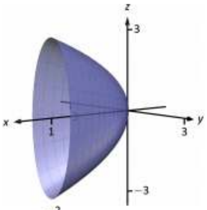
20. Revolve $y = x^2$ about the x-axis.

21. Revolve $z = x^2$ about the z-axis.

22. Revolve $z = 1/x$ about the z-axis.

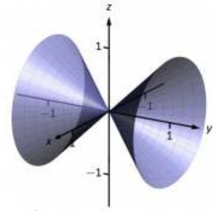
In Exercises 23-26, a quadric surface is sketched. Determine which of the given equations best fits the graph.

23.



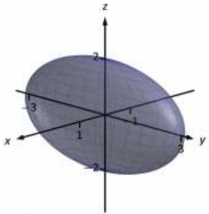
(a) $x=y^2+\frac{z^2}{9}$ (b) $x=y^2+\frac{z^2}{3}$

24.

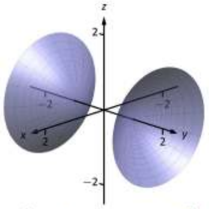


(a) $x^2-y^2-z^2=0$ (b) $x^2-y^2+z^2=0$

25.



(a) $x^2+\frac{y^2}{3}+\frac{z^2}{2}=1$ (b) $x^2+\frac{y^2}{9}+\frac{z^2}{4}=1$



(a) $y^2-x^2-z^2=1$ (b) $y^2+x^2-z^2=1$

26.

In Exercises 27-32, sketch the quadric surface.

27. $z-y^2+x^2=0$

28. $z^2=x^2+\frac{y^2}{4}$

29. $x=-y^2-z^2$

30. $16x^2-16y^2-16z^2=1$

31. $\frac{x^2}{9}-y^2+\frac{z^2}{25}=1$

32. $4x^2+2y^2+z^2=4$

10.2: An Introduction to Vectors

Terms and Concepts

1. Name two different things that cannot be described with just one number, but rather need 2 or more numbers to fully describe them.

2. What is the difference between $(1,2)$ and $\langle 1,2\rangle$?

3. What is a unit vector?

4. What does it mean for two vectors to be parallel?

5. What effect does multiplying by a vector by -2 have?

Problems

In Exercises 6-9, points *P* and *Q* are given. Write the vector $\vec{PQ}$ in component form and using the standard unit vectors.

6. $P=(2,-1), \quad Q=(3,5)$

7. $P=(3,2), \quad Q=(7,-2)$

8. $P=(0,3,-1), \quad Q=(6,2,5)$

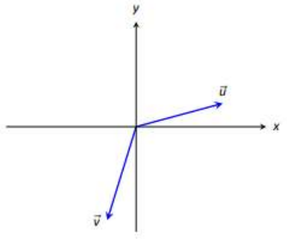
9. $P=(2,1,2), \quad Q=(4,3,2)$

10. Let $\vec{u}=\langle 1,-2\rangle$ and $\vec{v}=\langle 1,1\rangle$.
(a) Find $\vec{u}+\vec{v}$, $\vec{u}-\vec{v}$, $2\vec{u}-3\vec{v}$.
(b) Sketch the above vectors on the same axes, along with $\vec{u}$ and $\vec{v}$.
(c) Find $\vec{x}$ where $\vec{u}+\vec{x}=2\vec{v}-\vec{x}$.

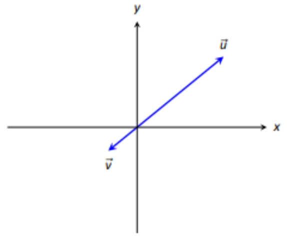
11. Let $\vec{u}=\langle 1,1,-1\rangle$ and $\vec{v}=\langle 2,1,2\rangle$.
(a) Find $\vec{u}+\vec{v}$, $\vec{u}-\vec{v}$, $\pi\vec{u}-\sqrt{2}\vec{v}$.
(b) Sketch the above vectors on the same axes, along with $\vec{u}$ and $\vec{v}$.
(c) Find $\vec{x}$ where $\vec{u}+\vec{x}=\vec{v}+2\vec{x}$.

In Exercises 12-15, sketch $\vec{u}, \vec{v}, \vec{u}+\vec{v}$ and $\vec{u}-\vec{v}$ on the same axes.

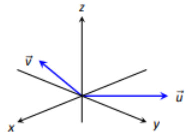
12.



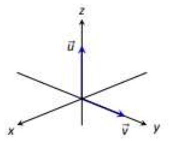
13.



14.



15.



In Exercises 16-19, find $\|\vec{u}\|, \|\vec{v}\|, \|\vec{u}+\vec{v}\|$ and $\|\vec{u}-\vec{v}\|$.

16. $\vec{u}=\langle 2,1\rangle, \quad \vec{v}=\langle 3,-2\rangle$

17. $\vec{u}=\langle -3,2,2\rangle, \quad \vec{v}=\langle 1,-1,1\rangle$

18. $\vec{u}=\langle 1,2\rangle, \quad \vec{v}=\langle -3,-6\rangle$

19. $\vec{u}=\langle 2,-3,6\rangle, \quad \vec{v}=\langle 10,-15,30\rangle$

20. Under what conditions is $\|\vec{u}\|+\|\vec{v}\|=\|\vec{u}+\vec{v}\|$?

In Exercises 21-24, find the unit vector $\vec{u}$ in the direction of $\vec{v}$.

21. $\vec{v}=\langle 3,7\rangle$

22. $\vec{v}=\langle 6,8\rangle$

23. $\vec{v}=\langle 1,-2,2\rangle$

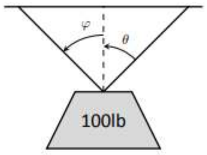
24. $\vec{v}=\langle 2,-2,2\rangle$

25. Find the unit vector in the first quadrant of $\mathbb{R}^2$ that makes a 50° angle with the x-axis.

26. Find the unit vector in the second quadrant of $\mathbb{R}^2$ that makes a 30° angle with the y-axis.

27. Verify, from Key Idea 48, that $\vec{u}=\langle \sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta\rangle$ is a unit vector for all angles θ and ϕ.

A weight of 100lb is suspended from two chains, making an angles with the vertical of θ and ϕ as shown in the figure below.



In Exercises 28-31, angles θ and ϕ are given. Find the force applied to each chain.

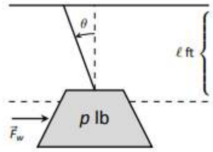
28. $\theta=30^\circ, \quad \phi=30^\circ$

29. $\theta=60^\circ, \quad \phi=60^\circ$

30. $\theta=20^\circ, \quad \phi=15^\circ$

31. $\theta=0^\circ, \quad \phi=0^\circ$

A weight of $p$lb is suspended from a chain of length l while a constant force of $\vec{F}_w$ pushes the weight to the right, making an angle of θ with the vertical, as shown in the figure below.



In Exercises 32-35, a force $\vec{F}_w$ and length l are given. Find the angle θ and the height the weight is lifted as it moves to the right.

32. $\vec{F}_w=1\text{lb}, \quad l=1\text{ft}, \quad p=1\text{lb}$

33. $\vec{F}_w = 1\text{lb}, \quad l = 1\text{ft}, \quad p = 10\text{lb}$

34. $\vec{F}_w = 1\text{lb}, \quad l = 10\text{ft}, \quad p = 1\text{lb}$

35. $\vec{F}_w = 10\text{lb}, \quad l = 10\text{ft}, \quad p = 1\text{lb}$

10.3: The Dot Product

Terms and Concepts

1. The dot product of two vectors is a __________, not a vector.

2. How are the concepts of the dot product and vector magnitude related?

3. How can one quickly tell if the angle between two vectors is acute or obtuse?

4. Give a synonym for "orthogonal."

Problems

In Exercises 5-10, find the dot product of the given vectors.

5. $\vec{u} = \langle 2, -4 \rangle, \vec{v} = \langle 3, 7 \rangle$

6. $\vec{u} = \langle 5, 3 \rangle, \vec{v} = \langle 6, 1 \rangle$

7. $\vec{u} = \langle 1, -1, 2 \rangle, \vec{v} = \langle 2, 5, 3 \rangle$

8. $\vec{u} = \langle 3, 5, -1 \rangle, \vec{v} = \langle 4, -1, 7 \rangle$

9. $\vec{u} = \langle 1, 1 \rangle, \vec{v} = \langle 1, 2, 3 \rangle$

10. $\vec{u} = \langle 1, 2, 3 \rangle, \vec{v} = \langle 0, 0, 0 \rangle$

11. Create your own vectors $\vec{u}, \vec{v}, \text{ and } \vec{w} \text{ in } \mathbb{R}^2$ and show that $\vec{u} \cdot (\vec{v} + \vec{w}) = \vec{u} \cdot \vec{v} + \vec{u} \cdot \vec{w}$.

12. Create your own vectors $\vec{u} \text{ and } \vec{v} \text{ in } \mathbb{R}^3$ and scalar *c* and show that $c(\vec{u} \cdot \vec{v}) = \vec{u} \cdot (c\vec{v})$.

In Exercises 13-16, find the measure of the angle between the two vectors in both radians and degrees.

13. $\vec{u} = \langle 1, 1 \rangle, \ \vec{v} = \langle 1, 2 \rangle$

14. $\vec{u} = \langle -2, 1 \rangle, \ \vec{v} = \langle 3, 5 \rangle$

15. $\vec{u} = \langle 8, 1, -4 \rangle, \ \vec{v} = \langle 2, 2, 0 \rangle$

16. $\vec{u} = \langle 1, 7, 2 \rangle, \ \vec{v} = \langle 4, -2, 5 \rangle$

In Exercises 17-20, a vectors $\vec{v}$ is given. Give two vectors that are orthogonal to $\vec{v}$.

17. $\vec{v} = \langle 4, 7 \rangle$

18. $\vec{v} = \langle -3, 5 \rangle$

19. $\vec{v} = \langle 1, 1, 1 \rangle$

20. $\vec{v} = \langle 1, -2, 3 \rangle$

In Exercises 21-26, vectors $\vec{u} \text{ and } \vec{v}$ are given. Find $\text{proj}_{\vec{v}}\vec{u}$, the orthogonal projection of $\vec{u}$ onto $\vec{v}$, and sketch all three vectors on the same axes.

21. $\vec{u} = \langle 1, 2 \rangle, \vec{v} = \langle -1, 3 \rangle$

22. $\vec{u} = \langle 5, 5 \rangle, \vec{v} = \langle 1, 3 \rangle$

23. $\vec{u} = \langle -3, 2 \rangle, \vec{v} = \langle 1, 1 \rangle$

24. $\vec{u} = \langle -3, 2 \rangle, \vec{v} = \langle 2, 3 \rangle$

25. $\vec{u} = \langle 1, 5, 1 \rangle, \vec{v} = \langle 1, 2, 3 \rangle$

26. $\vec{u} = \langle 3, -1, 2 \rangle, \vec{v} = \langle 2, 2, 1 \rangle$

In Exercises 27-32, vectors $\vec{u}$ and $vecv$ are given. Write $\vec{u}$ as the sum of two vectors, one of which is parallel to $\vec{v}$ and one of which is perpendicular to $\vec{v}$. Note: these are the same pairs of vectors as found in Exercises 21-26.

27. $\vec{u} = \langle 1, 2\rangle, \vec{v} = \langle -1, 3\rangle$

28. $\vec{u} = \langle 5, 5\rangle, \vec{v} = \langle 1, 3\rangle$

29. $\vec{u} = \langle -3, 2\rangle, \vec{v} = \langle 1, 1\rangle$

30. $\vec{u} = \langle -3, 2\rangle, \vec{v} = \langle 2, 3\rangle$

31. $\vec{u} = \langle 1, 5, 1\rangle, \vec{v} = \langle 1, 2, 3\rangle$

32. $\vec{u} = \langle 3, -1, 2\rangle, \vec{v} = \langle 2, 2, 1\rangle$

33. A 10lb box sits on a ramp that rises 4ft over a distance of 20ft. How much force is required to keep the box from sliding down the ramp?

34. A 10lb box sits on a 15ft ramp that makes a 30° angle with the horizontal. How much force is required to keep the box from sliding down the ramp?

35. How much work is performed in moving a box horizontally 10ft with a force of 20lb applied at an angle of 45° to the horizontal?

36. How much work is performed in moving a box horizontally 10ft with a force of 20lb applied at an angle of 10° to the horizontal?

37. How much work is performed in moving a box up the length of a ramp that rises 2ft over a distance of 10ft, with a force of 50lb applied horizontally?

38. How much work is performed in moving a box up the length of a ramp that rises 2ft over a distance of 10ft, with a force of 50lb applied at an angle of 45° to the horizontal?

39. How much work is performed in moving a box up the length of a 10ft ramp that makes a 5° angle with the horizontal, with 50lb of force applied in the direction of the ramp?

10.4: The Cross Product

Terms and Concepts

1. The cross product of two vectors is a ________, not a scalar.

2. One can visualize the direction of $\vec{u} \times \vec{v}$ using the _______ ________ ________.

3. Give synonym for "orthogonal."

4. T/F: A fundamental principle of the cross product is that $\vec{u} \times \vec{v}$ is orthogonal to $\vec{u}$ and $\vec{v}$.

5. _______ is a measure of the turning force applied to an object.

Problems

In Exercises 6-14, vectors $\vec{u}$ and $\vec{v}$ are given. Compute $\vec{u} \times \vec{v}$ and show this is orthogonal to both $\vec{u}$ and $\vec{v}$.

6. $\vec{u} = \langle 3, 2, -2\rangle,\ \vec{v} = \langle 0, 1, 5\rangle$

7. $\vec{u} = \langle 5, -4, 3\rangle,\ \vec{v} = \langle 2, -5, 1\rangle$

8. $\vec{u} = \langle 4, -5, -5\rangle,\ \vec{v} = \langle 3, 3, 4\rangle$

9. $\vec{u} = \langle -4, 7, -10\rangle,\ \vec{v} = \langle 4, 4, 1\rangle$

10. $\vec{u} = \langle 1, 0, 1\rangle,\ \vec{v} = \langle 5, 0, 7\rangle$

11. $\vec{u} = \langle 1, 5, -4\rangle,\ \vec{v} = \langle -2, -10, 8\rangle$

12. $\vec{u} = \vec{i},\ \vec{v} = \vec{j}$

13. $\vec{u} = \vec{i},\ \vec{v} = \vec{k}$

14. $\vec{u} = \vec{j},\ \vec{v} = \vec{j}$

15. Pick any vectors $\vec{u}, \vec{v}$ and $\vec{w}$ in $\mathbb{R}^3$ and show that $\vec{u} \times (\vec{v} + \vec{w} = \vec{u} \times \vec{v} + \vec{u} \times \vec{w}$.

16. Pick any vectors $\vec{u}, \vec{v}$ and $\vec{w}$ in $\mathbb{R}^3$ and show that $\vec{u} \cdot (\vec{v} \times \vec{w} = (\vec{u} \times \vec{v}) \cdot \vec{w}$.

In Exercises 17-20, the magnitude of vectors $\vec{u}$ and $\vec{v}$ in $\mathbb{R}^3$ are given, along with the angle θ between them. Use this information to find the magnitude of $\vec{u} \times \vec{v}$.

17. $\|\vec{u}\| = 2, \quad \|\vec{v}\| = 5, \quad \theta = 30^\circ$

18. $\|\vec{u}\| = 3, \quad \|\vec{v}\| = 7, \quad \theta = \pi/2$

19. $\|\vec{u}\| = 3, \quad \|\vec{v}\| = 4, \quad \theta = \pi$

20. $\|\vec{u}\| = 2, \quad \|\vec{v}\| = 5, \quad \theta = 5\pi/6$

In Exercises 21-24, find the area of the parallelogram defined by the given vectors.

21. $\vec{u} = \langle 1, 1, 2 \rangle$, $\vec{v} = \langle 2, 0, 3 \rangle$

22. $\vec{u} = \langle -2, 1, 5 \rangle$, $\vec{v} = \langle -1, 3, 1 \rangle$

23. $\vec{u} = \langle 1, 2 \rangle$, $\vec{v} = \langle 2, 1 \rangle$

24. $\vec{u} = \langle 2, 0 \rangle$, $\vec{v} = \langle 0, 3 \rangle$

In Exercises 25-28, find the area of the triangle with the given vertices.

25. Vertices: $(0, 0, 0), (1, 3, -1)$ and $(2, 1, 1)$.

26. Vertices: $(5, 2, -1), (3, 6, 2)$ and $(1, 0, 4)$.

27. Vertices: $(1, 1), (1, 3)$ and $(2, 2)$.

28. Vertices: $(3, 1), (1, 2)$ and $(4, 3)$.

In Exercises 29-30, find the area of the quadrilateral with the given vertices. (Hint: break the quadrilateral into the 2 triangles.)

29. Vertices: $(0, 0), (1, 2), (3, 0)$ and $(4, 3)$.

30. Vertices: $(0, 0, 0), (2, 1, 1), (-1, 2, -8)$ and $(1, -1, 5)$.

In Exercises 31-32, find the volume of the parallelepiped defined by the given vectors.

31. $\vec{u} = \langle 1, 1, 1 \rangle, \quad \vec{v} = \langle 1, 2, 3 \rangle, \quad \vec{w} = \langle 1, 0, 1 \rangle$

32. $\vec{u} = \langle -1, 2, 1 \rangle, \quad \vec{v} = \langle 2, 2, 1 \rangle, \quad \vec{w} = \langle 3, 1, 3 \rangle$

In Exercises 33-36, find a unit vector orthogonal to both $\vec{u}$ and $\vec{v}$.

33. $\vec{u} = \langle 1, 1, 1 \rangle, \quad \vec{v} = \langle 2, 0, 1 \rangle$

34. $\vec{u} = \langle 1, -2, 1 \rangle, \quad \vec{v} = \langle 3, 2, 1 \rangle$

35. $\vec{u} = \langle 5, 0, 2 \rangle, \quad \vec{v} = \langle -3, 0, 7 \rangle$

36. $\vec{u} = \langle 1, -2, 1 \rangle, \quad \vec{v} = \langle -2, 4, -2 \rangle$

37. A bicycle rider applies 150lb of force, straight down, onto a pedal that extends ȷin horizontally from the crankshaft. Find the magnitude of the torque applied to the crankshaft.

38. A bicycle rider applies 150lb of force, straight down, onto a pedal that extends 7in from the crankshaft, making a 30° angle with the horizontal. Find the magnitude of the torque applied to the crankshaft.

39. To turn a stubborn bolt, 80lb of force is applied to a 10in wrench. What is the maximum amount of torque that can be applied to the bolt?

40. To turn a stubborn bolt, 80lb of force is applied to a 10in wrench in a confined space, where the direction of applied force makes a 10° angle with the wrench. How much torque is subsequently applied to the wrench?

41. Show, using the definition of the Cross Product, that $\vec{u} \cdot (\vec{u} \times \vec{v}) = 0$; that is, that $\vec{u}$ is orthogonal to the cross product of $\vec{u}$ and $\vec{v}$.

42. Show, using the definition of the Cross Product, that $vecu \times \vec{u} = \vec{0}$.

10.5: Lines

Terms and Concepts

1. To find an equation of a line, what two pieces of information are needed?

2. Two distinct lines in the plane can intersect or be ________.

3. Two distinct lines in space can intersect, be __________ or be ___________.

4. Use your own words to describe what it means for two lines in space to be skew.

Problems

In Exercises 5-14, write the vector, parametric and symmetric equations of the lines described.

5. Passes through $P=(2,-4,1)$, parallel to $\vec{d}=\langle 9,2,5\rangle$.

6. Passes through $P=(6,1,7)$, parallel to $\vec{d}=\langle -3,2,5\rangle$.

7. Passes through $P=(2,1,5)$ and $Q=(7,-2,4)$.

8. Passes through $P=(1,-2,3)$ and $Q=(5,5,5)$.

9. Passes through $P=(0,1,2)$ and orthogonal to both $\vec{d}_1=\langle 2,-1,7\rangle$ and $\vec{d}_2=\langle 7,1,3\rangle$.

10. Passes through $P=(5,1,9)$ and orthogonal to both $\vec{d}_1=\langle 1,0,1\rangle$ and $\vec{d}_2=\langle 2,0,3\rangle$.

11. Passes through the point of intersection of $\vec{l}_1(t)$ and $\vec{l}_2(t)$ and orthogonal to both lines, where
$\vec{l}_1(t)=\langle 2,1,1\rangle+t\langle 5,1,-2\rangle$ and
$\vec{l}_2(t)=\langle -2,-1,2\rangle+t\langle 3,1,-1\rangle$.

12. Passes through the point of intersection of $\vec{l}_1(t)$ and $\vec{l}_2(t)$ and orthogonal to both lines, where

$$l_1=\begin{cases}x=t\\ y=-2+2t\\ z=1+t\end{cases} \text{ and } l_2=\begin{cases}x=2+t\\ y=2-t\\ z=1+t\end{cases}$$

13. Passes through $P=(1,1)$, parallel to $\vec{d}=\langle 2,3\rangle$.

14. Passes through $P=(-2,5)$, parallel to $\vec{d}=\langle 0,1\rangle$.

In Exercises 15-22, determine if the described lines are the same line, parallel lines, intersecting or skew lines. If intersecting, give the point of intersection.

15.
$\vec{l}_1(t)=\langle 1,2,1\rangle+t\langle 2,-1,1\rangle, \vec{l}_2(t)=\langle 3,3,3\rangle+t\langle -4,2,-2\rangle$.

16.
$\vec{l}_1(t)=\langle 2,1,1\rangle+t\langle 5,1,3\rangle, \vec{l}_2(t)=\langle 14,5,9\rangle+t\langle 1,1,1\rangle$.

17.
$\vec{l}_1(t)=\langle 3,4,1\rangle+t\langle 2,-3,4\rangle, \vec{l}_2(t)=\langle -3,3,-3\rangle+t\langle 3,-2,4\rangle$.

18.
$\vec{l}_1(t)=\langle 1,1,1\rangle+t\langle 3,1,3\rangle, \vec{l}_2(t)=\langle 7,3,7\rangle+t\langle 6,2,6\rangle$.

19.

$$l_1=\begin{cases}x=1+2t\\ y=3-2t\\ z=t\end{cases} \text{ and } l_2=\begin{cases}x=3-t\\ y=2+5t\\ z=2+7t\end{cases}$$

20.

$$l_1=\begin{cases}x=1.1+0.6t\\ y=3.77+0.9t\\ z=-2.3+1.5t\end{cases} \text{ and } l_2=\begin{cases}x=3.11+3.4t\\ y=2+5.1t\\ z=2.5+8.5t\end{cases}$$

21.

$$l_1=\begin{cases}x=-0.2+0.6t\\ y=1.33-0.45t\\ z=-4.2+1.05t\end{cases} \text{ and } l_2=\begin{cases}x=0.86+9.2t\\ y=0.835+9.2t\\ z=-3.045+16.1t\end{cases}$$

22.
$$l_1 = \begin{cases} x = 0.1 + 1.1t \\ y = 2.9 - 1.5t \\ z = 3.2 + 1.6t \end{cases} \text{ and } l_2 = \begin{cases} x = 4 - 2.1t \\ y = 1.8 + 7.2t \\ z = 3.1 + 1.1t \end{cases}$$

In Exercises 23-26, find the distance from the point to the line.

23. $P = (1,1,1), \quad \vec{l}\,(t) = \langle 2,1,3\rangle + t\langle 2,1,-2\rangle$

24. $P = (2,5,6), \quad \vec{l}\,(t) = \langle -1,1,1\rangle + t\langle 1,0,1\rangle$

25. $P = (0,3), \quad \vec{l}\,(t) = \langle 2,0\rangle + t\langle 1,1\rangle$

26. $P = (1,1), \quad \vec{l}\,(t) = \langle 4,5\rangle + t\langle -4,3\rangle$

In Exercises 27-28, find the distance between the two lines.

27.
$\vec{l}_1(t) = \langle 1,2,1\rangle + t\langle 2,-1,1\rangle, \vec{l}_2(t) = \langle 3,3,3\rangle + t\langle 4,2,-2\rangle.$

28.
$\vec{l}_1(t) = \langle 0,0,1\rangle + t\langle 1,0,0\rangle, \vec{l}_2(t) = \langle 0,0,3\rangle + t\langle 0,1,0\rangle.$

Exercises 29-31 explore special cases of the distance formulas found in Key Idea 50.

29. Let Q be a point on the line $l(t)$. Show why the distance formula correctly gives the distance from the point to the line as 0.

30. Let lines $l_1(t)$ and $l_2(t)$ be intersecting lines. Show why the distance formula correctly gives the distance between these lines as 0.

31. Let lines $l_1(t)$ and $l_2(t)$ be parallel.
(a) Show why the distance formula for distance between lines cannot be used as stated to find the distance between the lines.
(b) Show why letting $c = (\overrightarrow{P_1P_2} \times \vec{d}_2) \times \vec{d}_2$ allows one to the use the formula.
(c) Show how one can use the formula for the distance between a point and a line and a line to find the distance between parallel lines.

10.6: Planes

Terms and Concepts

1. In order to find the equation of a plane, what two pieces of information must one have?

2. What is the relationship between a plane and one of its normal vectors?

Problems

In Exercises 3-6, give any two points in the given plane.

3. $2x - 4y + 7z = 2$

4. $3(x+2) + 5(y-9) - 4z = 0$

5. $x = 2$

6. $4(y+2) - (z-6) = 0$

In Exercises 7-20, give the equation of the described plane in standard and general forms.

7. Passes through $(2,3,4)$ and has normal vector $\vec{n} = \langle 3,-1,7\rangle$.

8. Passes through $(1,3,5)$ and has normal vector $\vec{n} = \langle 0,2,4\rangle$.

9. Passes through the points $(1,2,3), (3,-1,4)$ and $(1,0,1)$.

10. Passes through the points $(5,3,8), (6,4,9)$ and $(3,3,3)$.

11. Contains the intersecting lines
$l_1(t) = \langle 2,1,2\rangle + t\langle 1,2,3\rangle$ and
$l_2(t) = \langle 2,1,2\rangle + t\langle 2,5,4\rangle$.

12. Contains the intersecting lines
$l_1(t) = \langle 5, 0, 3\rangle + t\langle -1, 1, 1\rangle$ and
$l_2(t) = \langle 1, 4, 7\rangle + t\langle 3, 0, -3\rangle$.

13. Contains the parallel lines
$l_1(t) = \langle 1, 1, 1\rangle + t\langle 1, 2, 3\rangle$ and
$l_2(t) = \langle 1, 1, 2\rangle + t\langle 1, 2, 3\rangle$.

14. Contains the parallel lines
$l_1(t) = \langle 2, 1, 2\rangle + t\langle 1, 2, 3\rangle$ and
$l_2(t) = \langle 2, 1, 2\rangle + t\langle 2, 5, 4\rangle$.

15. Contains the point $(2, -6, 1)$ and the line
$$l(t) = \begin{cases} x = 2 + 5t \\ y = 2 + 2t \\ z = -1 + 2t \end{cases}$$

16. Contains the point $(5, 7, 3)$ and the line
$$l(t) = \begin{cases} x = t \\ y = t \\ z = t \end{cases}$$

17. Contains the point $(5, 7, 3)$ and is orthogonal to the line $l(t) = \langle 4, 5, 6\rangle + t\langle 1, 1, 1\rangle$.

18. Contains the point $(4, 1, 1)$ and is orthogonal to the line
$$l(t) = \begin{cases} x = 4 + 4t \\ y = 1 + 1t \\ z = 1 + 1t \end{cases}$$

19. Contains the point $(-4, 7, 2)$ and is parallel to the plane $3(x-2) + 8(y+1) - 10z = 0$.

20. Contains the point $(1, 2, 3)$ and is parallel to the plane $x = 5$.

In Exercises 21-22, give the equation of the line that is the intersection of the given plane.

21.
$p1:\ 3(x-2) + (y-1) + 4z = 0,\ \text{and}$
$p2:\ 2(x-1) - 2(y+3) + 6(z-1) = 0.$

22.
$p1:\ 5(x-5) + 2(y+2) + 4(z-1) = 0,\ \text{and}$
$p2:\ 3x - 4(y-1) + 2(z-1) = 0.$

In Exercises 23-26, find the point of intersection between the line and the plane.

23.
line: $\langle 5, 1, -1\rangle + t\langle 2, 2, 1\rangle$,
plane: $5x - y - z = -3$

24.
line: $\langle 4, 1, 0\rangle + t\langle 1, 0, -1\rangle$,
plane: $3x + y - 2x = 8$

25.
line: $\langle 1, 2, 3\rangle + t\langle 3, 5, -1\rangle$,
plane: $3x - 2y - z = 4$

26.
line: $\langle 1, 2, 3\rangle + t\langle 3, 5, -1\rangle$,
plane: $3x - 2y - z = -4$

In Exercises 27-30, find the given distances.

27. The distance from the point $(1, 2, 3)$ to the plane
$3(x-1) + (y-2) + 5(z-2) = 0$.

28. The distance from the point $(2, 6, 2)$ to the plane
$2(x-1) - y + 4(z+1) = 0.$

29. The distance between the parallel planes
$x + y + z = 0$ and
$(x-2) + (y-3) + (z+4) = 0.$

30. The distance between parallel planes
$2(x-1) + 2(y+1) + (z-2) = 0$ and
$2(x-3) + 2(y-1) + (z-3) = 0$

31. Show why if the point Q lies in a plane, then the distance formula correctly gives the distance from the point to the plane as 0.

32. How is Exercise 30 in Section 10.5 easier to answer once we have an understanding of planes?

10.E: Applications of Vectors (Exercises) is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

11: Vector-Valued Functions

In the previous chapter, we learned about vectors and were introduced to the power of vectors within mathematics. In this chapter, we’ll build on this foundation to define functions whose input is a real number and whose output is a vector. We’ll see how to graph these functions and apply calculus techniques to analyze their behavior. Most importantly, we’ll see why we are interested in doing this: we’ll see beautiful applications to the study of moving objects.

11.1: Vector–Valued Functions
11.2: Calculus and Vector-Valued Functions
11.3: The Calculus of Motion
11.4: Unit Tangent and Normal Vectors
11.5: The Arc Length Parameter and Curvature
11.E: Applications of Vector Valued Functions (Exercises)

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 11: Vector-Valued Functions is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

11.1: Vector–Valued Functions

We are very familiar with **real valued functions**, that is, functions whose output is a real number. This section introduces **vector-valued functions** - functions whose output is a vector.

Definition 11.1.1: Vector-Valued Functions

A **vector-valued function** is a function of the form

$$\vec{\mathbf{r}}(t) = \langle\, f(t), g(t)\,\rangle \tag{11.1.1}$$

or

$$\vec{\mathbf{r}}(t) = \langle\, f(t), g(t), h(t)\,\rangle, \tag{11.1.2}$$

where f, g and h are real valued functions.

The **domain** of $\vec{\mathbf{r}}$ is the set of all values of t for which $\vec{\mathbf{r}}(t)$ is defined. The **range** of $\vec{\mathbf{r}}$ is the set of all possible output vectors $\vec{\mathbf{r}}(t)$.

Evaluating and Graphing Vector-Valued Functions

Evaluating a vector-valued function at a specific value of t is straightforward; simply evaluate each component function at that value of t. For instance, if $\vec{\mathbf{r}}(t) = \langle t^2, t^2 + t - 1\rangle$, then $\vec{\mathbf{r}}(-2) = \langle 4, 1\rangle$. We can sketch this vector, as is done in Figure $11.1.1a$. Plotting lots of vectors is cumbersome, though, so generally we do not sketch the whole vector but just the terminal point. The **graph** of a vector-valued function is the set of all terminal points of $\vec{\mathbf{r}}(t)$, where the initial point of each vector is always the origin. In Figure $11.1.1b$ we sketch the graph of $\vec{\mathbf{r}}$; we can indicate individual points on the graph with their respective vector, as shown.

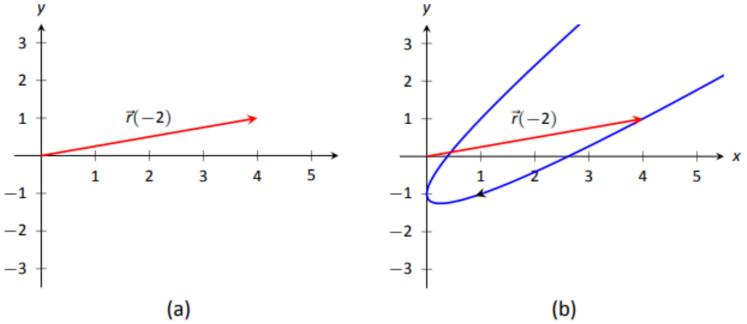


Figure $11.1.1$: Sketching the graph of a vector-valued function.

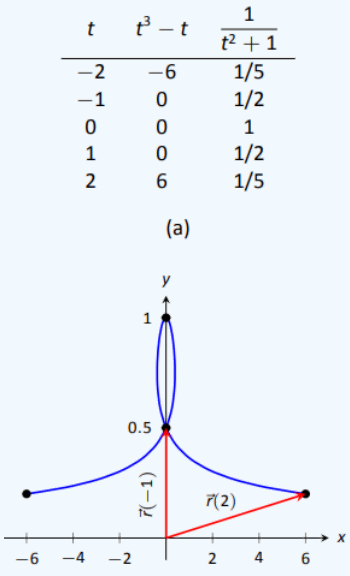
Vector-valued functions are closely related to parametric equations of graphs. While in both methods we plot points $\big(x(t), y(t)\big)$ or $\big(x(t), y(t), z(t)\big)$ to produce a graph, in the context of vector-valued functions each such point represents a vector. The implications of this will be more fully realized in the next section as we apply calculus ideas to these functions.

Example 11.1.1: Graphing vector-valued functions

Graph $\vec{\mathbf{r}}(t) = \langle t^3 - t, \dfrac{1}{t^2+1}\rangle$, for $-2 \le t \le 2$. Sketch $\vec{\mathbf{r}}(-1)$ and $\vec{\mathbf{r}}(2)$.

Solution

We start by making a table of t, x and y values as shown in Figure $11.1.1a$. Plotting these points gives an indication of what the graph looks like. In Figure $11.1.1b$, we indicate these points and sketch the full graph. We also highlight $\vec{\mathbf{r}}(-1)$ and $\vec{\mathbf{r}}(2)$ on the graph.



t	$t^3 - t$	$\frac{1}{t^2+1}$
-2	-6	$1/5$
-1	0	$1/2$
0	0	1
1	0	$1/2$
2	6	$1/5$

(a)

(b)

Figure 11.1.2: Sketching the vector-valued function of Example 11.1.1

Example 11.1.2: Graphing vector-valued functions.

Graph $\vec{\mathbf{r}}(t) = \langle \cos t, \sin t, t \rangle$ for $0 \leq t \leq 4\pi$.

Solution

We can again plot points, but careful consideration of this function is very revealing. Momentarily ignoring the third component, we see the x and y components trace out a circle of radius 1 centered at the origin. Noticing that the z component is t, we see that as the graph winds around the z-axis, it is also increasing at a constant rate in the positive z direction, forming a spiral. This is graphed in Figure 11.1.3. In the graph $\vec{\mathbf{r}}(7\pi/4) \approx (0.707, -0.707, 5.498)$ is highlighted to help us understand the graph.

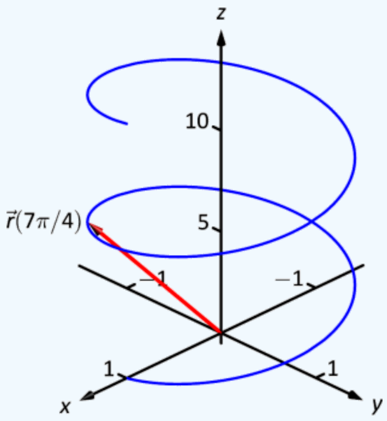


Figure 11.1.3: Viewing a vector-valued function, and its derivative at one point.

Algebra of Vector-Valued Functions

Definition 11.1.2: Operations on Vector-Valued Functions

Let $\vec{\mathbf{r}}_1(t) = \langle f_1(t), g_1(t) \rangle$ and $\vec{\mathbf{r}}_2(t) = \langle f_2(t), g_2(t) \rangle$ be vector-valued functions in $\mathbb{R}^2$ and let c be a scalar. Then:

1. $\vec{\mathbf{r}}_1(t) \pm \vec{\mathbf{r}}_2(t) = \langle\, f_1(t) \pm f_2(t), g_1(t) \pm g_2(t) \,\rangle$.
2. $c\vec{\mathbf{r}}_1(t) = \langle\, cf_1(t), cg_1(t) \,\rangle$.

A similar definition holds for vector-valued functions in $\mathbb{R}^3$.

This definition states that we add, subtract and scale vector-valued functions component-wise. Combining vector-valued functions in this way can be very useful (as well as create interesting graphs).

Example 11.1.3: Adding and scaling vector-valued functions.

Let $\vec{r}_1(t) = \langle 0.2t, 0.3t\rangle$, $\vec{r}_2(t) = \langle \cos t, \sin t\rangle$ and $\vec{r}(t) = \vec{r}_1(t) + \vec{r}_2(t)$. Graph $\vec{r}_1(t)$, $\vec{r}_2(t)$, $\vec{r}(t)$ and $5\vec{r}(t)$ on $-10 \le t \le 10$.

Solution

We can graph $\vec{r}_1$ and $\vec{r}_2$ easily by plotting points (or just using technology). Let's think about each for a moment to better understand how vector-valued functions work.

We can rewrite $\vec{r}_1(t) = \langle 0.2t, 0.3t\rangle$ as $\vec{r}_1(t) = t\langle 0.2, 0.3\rangle$. That is, the function $\vec{r}_1$ scales the vector $\langle 0.2, 0.3\rangle$ by t. This scaling of a vector produces a line in the direction of $\langle 0.2, 0.3\rangle$.

We are familiar with $\vec{r}_2(t) = \langle \cos t, \sin t\rangle$; it traces out a circle, centered at the origin, of radius 1. Figure $11.1.4a$ graphs $\vec{r}_1(t)$ and $\vec{r}_2(t)$.

Adding $\vec{r}_1(t)$ to $\vec{r}_2(t)$ produces $\vec{r}(t) = \langle \cos t + 0.2t, \sin t + 0.3t\rangle$, graphed in Figure $11.1.4b$. The linear movement of the line combines with the circle to create loops that move in the direction of $\langle 0.2, 0.3\rangle$. (We encourage the reader to experiment by changing $\vec{r}_1(t)$ to $\langle 2t, 3t\rangle$, etc., and observe the effects on the loops.)

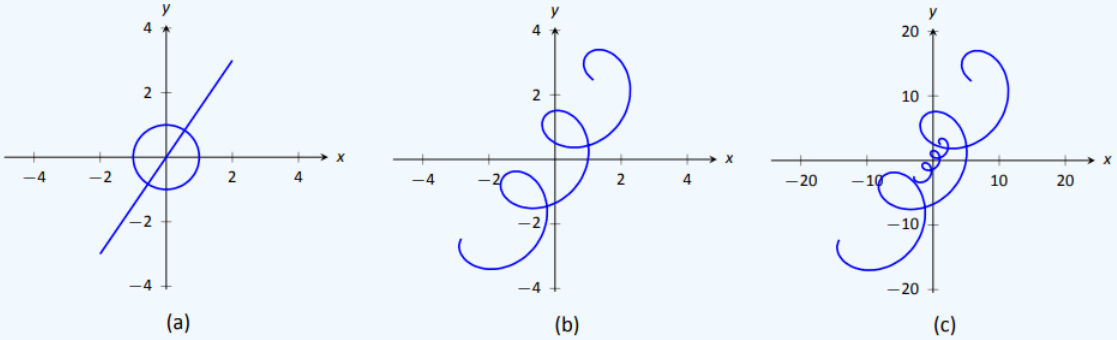


Figure 11.1.4: Graphing the functions in Example 11.1.3

Multiplying $\vec{r}(t)$ by 5 scales the function by 5, producing $5\vec{r}(t) = \langle 5\cos t + 1, 5\sin t + 1.5\rangle$, which is graphed in Figure $11.1.4c$ along with $\vec{r}(t)$. The new function is "5 times bigger" than $\vec{r}(t)$. Note how the graph of $5\vec{r}(t)$ in (c) looks identical to the graph of $\vec{r}(t)$ in (b). This is due to the fact that the x and y bounds of the plot in (c) are exactly 5 times larger than the bounds in (b).

Example 11.1.4: Adding and scaling vector-valued functions.

A **cycloid** is a graph traced by a point p on a rolling circle, as shown in Figure 11.1.5. Find an equation describing the cycloid, where the circle has radius 1.

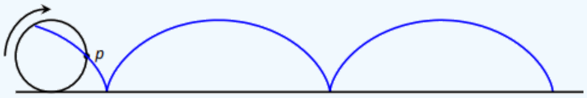


Figure 11.1.5: Tracing a cycloid.

Solution

This problem is not very difficult if we approach it in a clever way. We start by letting $\vec{p}(t)$ describe the position of the point p on the circle, where the circle is centered at the origin and only rotates clockwise (i.e., it does not roll). This is relatively simple given our previous experiences with parametric equations; $\vec{p}(t) = \langle \cos t, -\sin t\rangle$.

We now want the circle to roll. We represent this by letting $\vec{c}(t)$ represent the location of the center of the circle. It should be clear that the y component of $\vec{c}(t)$ should be 1; the center of the circle is always going to be 1 if it rolls on a horizontal surface.

The x component of $\vec{c}(t)$ is a linear function of t: $f(t) = mt$ for some scalar m. When $t = 0$, $f(t) = 0$ (the circle starts centered on the y-axis). When $t = 2\pi$, the circle has made one complete revolution, traveling a distance equal to its circumference, which is also 2π. This gives us a point on our line $f(t) = mt$, the point $(2\pi, 2\pi)$. It should be clear that $m = 1$ and $f(t) = t$. So $\vec{c}(t) = \langle t, 1\rangle$.

We now combine $\vec{\mathbf{p}}$ and $\vec{\mathbf{c}}$ together to form the equation of the cycloid:

$$\vec{\mathbf{r}}(t)=\vec{\mathbf{p}}(t)+\vec{\mathbf{c}}(t)=\langle\cos t+t,-\sin t+1\rangle,$$

which is graphed in Figure 11.1.6.

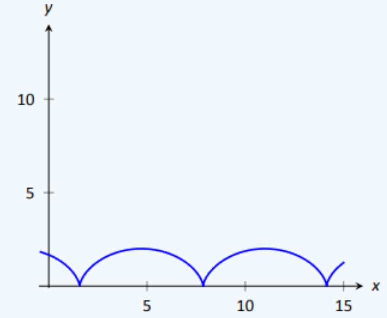


Figure 11.1.6: The cycloid in Example 11.1.4.

Displacement

A vector-valued function $\vec{\mathbf{r}}(t)$ is often used to describe the position of a moving object at time t. At $t=t_0$, the object is at $\vec{\mathbf{r}}(t_0)$; at $t=t_1$, the object is at $\vec{\mathbf{r}}(t_1)$. Knowing the locations $\vec{\mathbf{r}}(t_0)$ and $\vec{\mathbf{r}}(t_1)$ give no indication of the path taken between them, but often we only care about the difference of the locations, $\vec{\mathbf{r}}(t_1)-\vec{\mathbf{r}}(t_0)$, the *displacement*.

Definition 11.1.3: Displacement

Let $\vec{\mathbf{r}}(t)$ be a vector-valued function and let $t_0<t_1$ be values in the domain. The *displacement* $\vec{\mathbf{d}}$ of $\vec{\mathbf{r}}$, from $t=t_0$ to $t=t_1$, is

$$\vec{\mathbf{d}}=\vec{\mathbf{r}}(t_1)-\vec{\mathbf{r}}(t_0). \tag{11.1.3}$$

When the displacement vector is drawn with initial point at $\vec{\mathbf{r}}(t_0)$, its terminal point is $\vec{\mathbf{r}}(t_1)$. We think of it as the vector which points from a starting position to an ending position.

Example 11.1.5: Finding and graphing displacement vectors

Let $\vec{\mathbf{r}}(t)=\langle\cos(\frac{\pi}{2}t),\sin(\frac{\pi}{2}t)\rangle$. Graph $\vec{\mathbf{r}}(t)$ on $-1\le t\le 1$, and find the displacement of $\vec{\mathbf{r}}(t)$ on this interval.

Solution

The function $\vec{\mathbf{r}}(t)$ traces out the unit circle, though at a different rate than the "usual" $\langle\cos t,\sin t\rangle$ parametrization. At $t_0=-1$, we have $\vec{\mathbf{r}}(t_0)=\langle 0,-1\rangle$; at $t_1=1$, we have $\vec{\mathbf{r}}(t_1)=\langle 0,1\rangle$. The displacement of $\vec{\mathbf{r}}(t)$ on $[-1,1]$ is thus

$$\vec{\mathbf{d}}=\langle 0,1\rangle-\langle 0,-1\rangle=\langle 0,2\rangle.$$

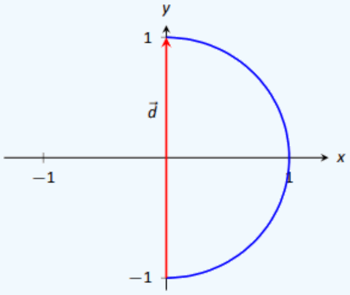


Figure 11.1.7: Graphing the displacement of a position function in Example 11.1.5.

A graph of $\vec{\mathbf{r}}(t)$ on $[-1,1]$ is given in Figure 11.1.7, along with the displacement vector $\vec{\mathbf{d}}$ on this interval.

Measuring displacement makes us contemplate related, yet very different, concepts. Considering the semi-circular path the object in Example 11.1.5 took, we can quickly verify that the object ended up a distance of 2 units from its initial location. That is, we can compute $\|d\|=2$. However, measuring *distance from the starting point* is different from measuring **distance traveled**. Being a semi-circle, we can measure the distance traveled by this object as $\pi \approx 3.14$ units. Knowing *distance from the starting point* allows us to compute **average rate of change**.

Definition 11.1.4: Average Rate of Change

Let $\vec{\mathbf{r}}(t)$ be a vector-valued function, where each of its component functions is continuous on its domain, and let $t_0 < t_1$. The **average rate of change** of $\vec{\mathbf{r}}(t)$ on $[t_0, t_1]$ is

$$\text{average rate of change} = \frac{\vec{\mathbf{r}}(t_1) - \vec{\mathbf{r}}(t_0)}{t_1 - t_0}. \tag{11.1.4}$$

Example 11.1.6: Average rate of change

Let $\vec{\mathbf{r}}(t) = \langle \cos(\frac{\pi}{2}t), \sin(\frac{\pi}{2}t)\rangle$ as in Example 11.1.5. Find the average rate of change of $\vec{\mathbf{r}}(t)$ on $[-1, 1]$ and on $[-1, 5]$.

Solution

We computed in Example 11.1.5 that the displacement of $\vec{\mathbf{r}}(t)$ on $[-1, 1]$ was $\vec{\mathbf{d}} = \langle 0, 2\rangle$. Thus the average rate of change of $\vec{\mathbf{r}}(t)$ on $[-1, 1]$ is:

$$\frac{\vec{\mathbf{r}}(1) - \vec{\mathbf{r}}(-1)}{1 - (-1)} = \frac{\langle 0, 2\rangle}{2} = \langle 0, 1\rangle.$$

We interpret this as follows: the object followed a semi-circular path, meaning it moved towards the right then moved back to the left, while climbing slowly, then quickly, then slowly again. *On average*, however, it progressed straight up at a constant rate of $\langle 0, 1\rangle$ per unit of time.

We can quickly see that the displacement on $[-1, 5]$ is the same as on $[-1, 1]$, so $\vec{\mathbf{d}} = \langle 0, 2\rangle$. The average rate of change is different, though:

$$\frac{\vec{\mathbf{r}}(5) - \vec{\mathbf{r}}(-1)}{5 - (-1)} = \frac{\langle 0, 2\rangle}{6} = \langle 0, 1/3\rangle.$$

As it took "3 times as long" to arrive at the same place, this average rate of change on $[-1, 5]$ is $1/3$ the average rate of change on $[-1, 1]$.

We considered average rates of change in Sections 1.1 and 2.1 as we studied limits and derivatives. The same is true here; in the following section we apply calculus concepts to vector-valued functions as we find limits, derivatives, and integrals. Understanding the average rate of change will give us an understanding of the derivative; displacement gives us one application of integration.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 11.1: Vector–Valued Functions is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

11.2: Calculus and Vector-Valued Functions

The previous section introduced us to a new mathematical object, the vector-valued function. We now apply calculus concepts to these functions. We start with the limit, then work our way through derivatives to integrals.

Limits of Vector-Valued Functions

The initial definition of the limit of a vector-valued function is a bit intimidating, as was the definition of the limit in Definition 1. The theorem following the definition shows that in practice, taking limits of vector-valued functions is no more difficult than taking limits of real-valued functions.

Definition 68: Limits of Vector-Valued Functions

Let I be an open interval containing c, and let $\vec{r}(t)$ be a vector-valued function defined on I, except possibly at c. The **limit of** $\vec{r}(t)$**, as** t **approaches** c**, is** $\vec{L}$, expressed as

$$\lim_{t\to c}\vec{r}(t)=\vec{L}, \tag{11.2.1}$$

means that given any $\epsilon>0$, there exists a $\delta>0$ such that for all $t\neq c$, if $|t-c|<\delta$, we have $\|\vec{r}(t)-\vec{L}\|<\epsilon$.

Note how the measurement of distance between real numbers is the absolute value of their difference; the measure of distance between vectors is the vector norm, or magnitude, of their difference.

theorem 89: Limits of Vector-Valued Functions

1. Let $\vec{r}(t)=\langle\, f(t),g(t)\,\rangle$ be a vector-valued function in $\mathbb{R}^2$ defined on an open interval I containing c. Then

$$\lim_{t\to c}\vec{r}(t)=\langle\lim_{t\to c}f(t)\,,\,\lim_{t\to c}g(t)\rangle. \tag{11.2.2}$$

2. Let $\vec{r}(t)=\langle\, f(t),g(t),h(t)\,\rangle$ be a vector-valued function in $\mathbb{R}^3$ defined on an open interval I containing c. Then

$$\lim_{t\to c}\vec{r}(t)=\langle\lim_{t\to c}f(t)\,,\,\lim_{t\to c}g(t)\,,\,\lim_{t\to c}h(t)\rangle \tag{11.2.3}$$

Theorem 89 states that we compute limits component-wise.

Example 11.2.1: Finding limits of vector-valued functions

Let $\vec{r}(t)=\langle\frac{\sin t}{t},\ t^2-3t+3,\ \cos t\rangle$. Find $\lim_{t\to 0}\vec{r}(t)$.

Solution

We apply the theorem and compute limits component-wise.

$$\begin{aligned}\lim_{t\to 0}\vec{r}(t)&=\langle\lim_{t\to 0}\frac{\sin t}{t}\,,\,\lim_{t\to 0}t^2-3t+3\,,\,\lim_{t\to 0}\cos t\rangle\\&=\langle 1,3,1\rangle.\end{aligned}$$

Continuity

Definition 69: Continuity of Vector-Valued Functions

Let $\vec{r}(t)$ be a vector-valued function defined on an open interval I containing c.

1. $\vec{r}(t)$ is **continuous at** c if $\lim_{t\to c}\vec{r}(t)=r(c)$.
2. If $\vec{r}(t)$ is continuous at all c in I, then $\vec{r}(t)$ is **continuous on** I.

We again have a theorem that lets us evaluate continuity component-wise.

THEOREM 90: Continuity of Vector-Valued Functions

Let $\vec{\mathbf{r}}(t)$ be a vector-valued function defined on an open interval I containing c. $\vec{\mathbf{r}}(t)$ is continuous at c if, and only if, each of its component functions is continuous at c.

Example 11.2.2: Evaluating continuity of vector-valued functions

Let $\vec{\mathbf{r}}(t) = \langle \frac{\sin t}{t},\, t^2 - 3t + 3,\, \cos t \rangle$. Determine whether $\vec{\mathbf{r}}$ is continuous at $t = 0$ and $t = 1$.

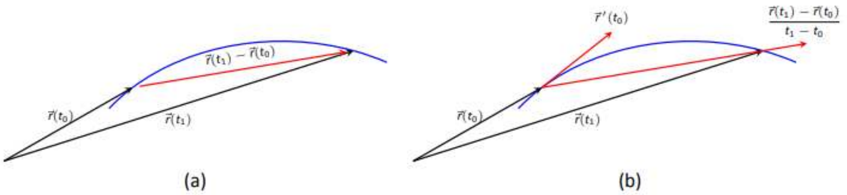
Solution

While the second and third components of $\vec{\mathbf{r}}(t)$ are defined at $t = 0$, the first component, $(\sin t)/t$, is not. Since the first component is not even defined at $t = 0$, $\vec{\mathbf{r}}(t)$ is not defined at $t = 0$, and hence it is not continuous at $t = 0$.

At $t = 1$ each of the component functions is continuous. Therefore $\vec{\mathbf{r}}(t)$ is continuous at $t = 1$.

Derivatives

Consider a vector-valued function $\vec{\mathbf{r}}$ defined on an open interval I containing t_0 and t_1. We can compute the displacement of $\vec{\mathbf{r}}$ on $[t_0, t_1]$, as shown in Figure $11.2.1a$. Recall that dividing the displacement vector by $t_1 - t_0$ gives the average rate of change on $[t_0, t_1]$, as shown in $11.2.1b$.


$\vec{r}(t_0)$
$\vec{r}(t_1)$
$\vec{r}(t_1) - \vec{r}(t_0)$
$\vec{r}\,'(t_0)$
$\vec{r}(t_0)$
$\vec{r}(t_1)$
$\frac{\vec{r}(t_1) - \vec{r}(t_0)}{t_1 - t_0}$

(a)

(b)

Figure $11.2.1$: Illustrating displacement, leading to an understanding of the derivative of vector-valued functions.

The **derivative** of a vector-valued function is a measure of the *instantaneous* rate of change, measured by taking the limit as the length of $[t_0, t_1]$ goes to 0. Instead of thinking of an interval as $[t_0, t_1]$, we think of it as $[c, c+h]$ for some value of h (hence the interval has length h). The *average* rate of change is

$$\frac{\vec{\mathbf{r}}(c+h) - \vec{\mathbf{r}}(c)}{h} \tag{11.2.4}$$

for any value of $h \neq 0$. We take the limit as $h \to 0$ to measure the instantaneous rate of change; this is the derivative of $\vec{\mathbf{r}}$.

Definition 70: Derivative of a Vector-Valued Function

Let $\vec{\mathbf{r}}(t)$ be continuous on an open interval I containing c.

1. The derivative of $\vec{\mathbf{r}}$ at $t = c$ is

$$\vec{\mathbf{r}}\,'(c) = \lim_{h\to 0} \frac{\vec{\mathbf{r}}(c+h) - \vec{\mathbf{r}}(c)}{h}. \tag{11.2.5}$$

2. The derivative of $\vec{\mathbf{r}}$ is

$$\vec{\mathbf{r}}\,'(t) = \lim_{h\to 0} \frac{\vec{\mathbf{r}}(t+h) - \vec{\mathbf{r}}(t)}{h}. \tag{11.2.6}$$

Note: Alternate notations for the derivative of $\vec{\mathbf{r}}$ include:

$$\vec{\mathbf{r}}\,'(t) = \frac{d}{dt}\Big(\vec{\mathbf{r}}(t)\Big) = \frac{d\vec{\mathbf{r}}}{dt}. \tag{11.2.7}$$

If a vector-valued function has a derivative for all c in an open interval I, we say that $\vec{\mathbf{r}}(t)$ is **differentiable** on I.

Once again we might view this definition as intimidating, but recall that we can evaluate limits component-wise. The following theorem verifies that this means we can compute derivatives component-wise as well, making the task not too difficult.

theorem 91: Derivatives of Vector-Valued Functions

1. Let $\vec{\mathbf{r}}(t) = \langle\, f(t), g(t)\,\rangle$. Then

$$\vec{\mathbf{r}}\,'(t) = \langle\, f'(t), g'(t)\,\rangle. \tag{11.2.8}$$

2. Let $\vec{\mathbf{r}}(t) = \langle\, f(t), g(t), h(t)\,\rangle$. Then

$$\vec{\mathbf{r}}\,'(t) = \langle\, f'(t), g'(t), h'(t)\,\rangle. \tag{11.2.9}$$

Example 11.2.3: Derivatives of vector-valued functions

Let $\vec{\mathbf{r}}(t) = \langle t^2, t\rangle$.

1. Sketch $\vec{\mathbf{r}}(t)$ and $\vec{\mathbf{r}}\,'(t)$ on the same axes.
2. Compute $\vec{\mathbf{r}}\,'(1)$ and sketch this vector with its initial point at the origin and at $\vec{\mathbf{r}}(1)$.

Solution

1. Theorem 91 allows us to compute derivatives component-wise, so

$$\vec{\mathbf{r}}\,'(t) = \langle 2t, 1\rangle. \tag{11.2.10}$$

   $\vec{\mathbf{r}}(t)$ and $\vec{\mathbf{r}}\,'(t)$ are graphed together in Figure 11.9(a). Note how plotting the two of these together, in this way, is not very illuminating. When dealing with real-valued functions, plotting $f(x)$ with $f'(x)$ gave us useful information as we were able to compare f and f' at the same x-values. When dealing with vector-valued functions, it is hard to tell which points on the graph of $\vec{\mathbf{r}}\,'$ correspond to which points on the graph of $\vec{\mathbf{r}}$.
2. We easily compute $\vec{\mathbf{r}}\,'(1) = \langle 2, 1\rangle$, which is drawn in Figure $11.2.2a$ with its initial point at the origin, as well as at $\vec{\mathbf{r}}(1) = \langle 1, 1\rangle$. These are sketched in Figure $11.2.2b$.

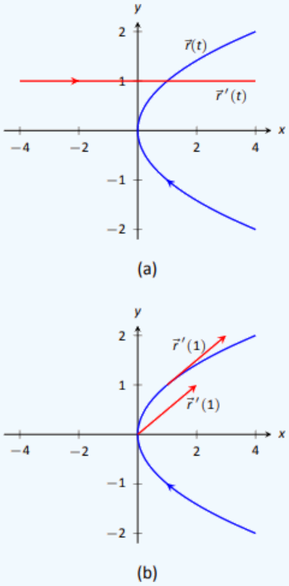


Figure $11.2.2$: Graphing the derivative of a vector-valued function in Example 11.2.3

Example 11.2.4: Derivatives of vector-valued functions

Let $\vec{\mathbf{r}}(t) = \langle \cos t, \sin t, t\rangle$. Compute $\vec{\mathbf{r}}\,'(t)$ and $\vec{\mathbf{r}}\,'(\pi/2)$. Sketch $\vec{\mathbf{r}}\,'(\pi/2)$ with its initial point at the origin and at $\vec{\mathbf{r}}(\pi/2)$.

Solution

We compute $\vec{\mathbf{r}}\,'$ as $\vec{\mathbf{r}}\,'(t) = \langle -\sin t, \cos t, 1\rangle$. At $t = \pi/2$, we have $\vec{\mathbf{r}}\,'(\pi/2) = \langle -1, 0, 1\rangle$. Figure $11.2.3$ shows a graph of $\vec{\mathbf{r}}(t)$, with $\vec{\mathbf{r}}\,'(\pi/2)$ plotted with its initial point at the origin and at $\vec{\mathbf{r}}(\pi/2)$.

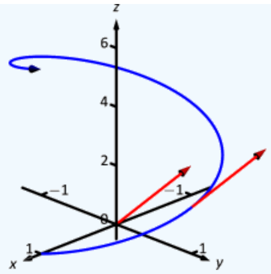


Figure 11.2.3: Viewing a vector-valued function and its derivative at one point.

In Examples 11.2.3 and 11.2.4, sketching a particular derivative with its initial point at the origin did not seem to reveal anything significant. However, when we sketched the vector with its initial point on the corresponding point on the graph, we did see something significant: the vector appeared to be *tangent* to the graph. We have not yet defined what "tangent" means in terms of curves in space; in fact, we use the derivative to define this term.

Definition 71: Tangent Vector, Tangent Line

Let $\vec{\mathbf{r}}(t)$ be a differentiable vector-valued function on an open interval I containing c, where $\vec{\mathbf{r}}'(c) \neq \vec{\mathbf{0}}$.

1. A vector $\vec{\mathbf{v}}$ is **tangent to the graph of** $\vec{\mathbf{r}}(t)$ at $t = c$ if $\vec{\mathbf{v}}$ is parallel to $\vec{\mathbf{r}}'(c)$.
2. The tangent line to the graph of $\vec{\mathbf{r}}(t)$ at $t = c$ is the line through $\vec{\mathbf{r}}(c)$ with direction parallel to $\vec{\mathbf{r}}'(c)$. An equation of the tangent line is

$$\vec{\ell}(t) = \vec{\mathbf{r}}(c) + t\,\vec{\mathbf{r}}'(c). \tag{11.2.11}$$

Example 11.2.5: Finding tangent lines to curves in space

Let $\vec{\mathbf{r}}(t) = \langle t, t^2, t^3 \rangle$ on $[-1.5, 1.5]$. Find the vector equation of the line tangent to the graph of $\vec{\mathbf{r}}$ at $t = -1$.

Solution

To find the equation of a line, we need a point on the line and the line's direction. The point is given by $\vec{\mathbf{r}}(-1) = \langle -1, 1, -1 \rangle$. (To be clear, $\langle -1, 1, -1 \rangle$ is a *vector*, not a point, but we use the point "pointed to" by this vector.)

The direction comes from $\vec{\mathbf{r}}'(-1)$. We compute, component-wise, $\vec{\mathbf{r}}'(t) = \langle 1, 2t, 3t^2 \rangle$. Thus $\vec{\mathbf{r}}'(-1) = \langle 1, -2, 3 \rangle$.

The vector equation of the line is $\ell(t) = \langle -1, 1, -1 \rangle + t\langle 1, -2, 3 \rangle$. This line and $\vec{\mathbf{r}}(t)$ are sketched, from two perspectives, in Figure 11.2.4a and 11.2.4b.

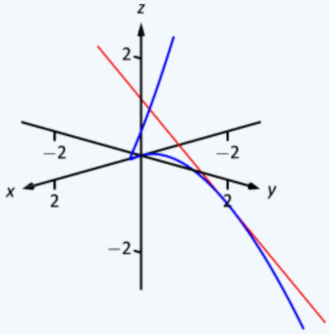


Figure 11.2.4: Graphing a curve in space with its tangent line.

Example 11.2.6: Finding tangent lines to curves

Find the equations of the lines tangent to $\vec{\mathbf{r}}(t) = \langle t^3, t^2 \rangle$ at $t = -1$ and $t = 0$.

Solution

We find that $\vec{\mathbf{r}}'(t) = \langle 3t^2, 2t \rangle$. At $t = -1$, we have

$$\vec{\mathbf{r}}(-1) = \langle -1, 1 \rangle \quad \text{and} \quad \vec{\mathbf{r}}'(-1) = \langle 3, -2 \rangle,$$

so the equation of the line tangent to the graph of $\vec{r}(t)$ at $t=-1$ is

$$\ell(t)=\langle -1,1\rangle+t\langle 3,-2\rangle.$$

This line is graphed with $\vec{r}(t)$ in Figure 11.2.5.

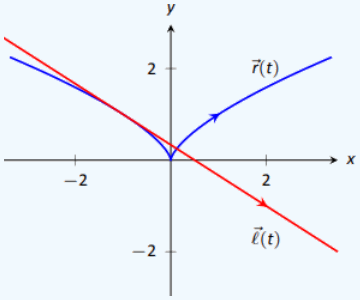


Figure 11.2.5: Graphing $\vec{r}(t)$ and its tangent line in Example 11.2.6.

At $t=0$, we have $\vec{r}\,'(0)=\langle 0,0\rangle=\vec{0}$! This implies that the tangent line "has no direction." We cannot apply Definition 71, hence cannot find the equation of the tangent line.

We were unable to compute the equation of the tangent line to $\vec{r}(t)=\langle t^3,t^2\rangle$ at $t=0$ because $\vec{r}\,'(0)=\vec{0}$. The graph in Figure 11.12 shows that there is a cusp at this point. This leads us to another definition of **smooth**, previously defined by Definition 46 in Section 9.2.

Definition 72: Smooth Vector-Valued Functions

Let $\vec{r}(t)$ be a differentiable vector-valued function on an open interval I. $\vec{r}(t)$ is **smooth** on I if $\vec{r}\,'(t)\neq\vec{0}$ on I.

Having established derivatives of vector-valued functions, we now explore the relationships between the derivative and other vector operations. The following theorem states how the derivative interacts with vector addition and the various vector products.

THEOREM 92: Properties of Derivatives of Vector-Valued Functions

Let $\vec{r}$ and $\vec{s}$ be differentiable vector-valued functions, let f be a differentiable real-valued function, and let c be a real number.

1. $\frac{d}{dt}\Big(\vec{r}(t)\pm\vec{s}(t)\Big)=\vec{r}\,'(t)\pm\vec{s}\,'(t)$
2. $\frac{d}{dt}\Big(c\vec{r}(t)\Big)=c\vec{r}\,'(t)$
3. $\frac{d}{dt}\Big(f(t)\vec{r}(t)\Big)=f'(t)\vec{r}(t)+f(t)\vec{r}\,'(t)$ **Product Rule**
4. $\frac{d}{dt}\Big(\vec{r}(t)\cdot\vec{s}(t)\Big)=\vec{r}\,'(t)\cdot\vec{s}(t)+\vec{r}(t)\cdot\vec{s}\,'(t)$ **Product Rule**
5. $\frac{d}{dt}\Big(\vec{r}(t)\times\vec{s}(t)\Big)=\vec{r}\,'(t)\times\vec{s}(t)+\vec{r}(t)\times\vec{s}\,'(t)$ **Product Rule**
6. $\frac{d}{dt}\Big(\vec{r}\big(f(t)\big)\Big)=\vec{r}\,'\big(f(t)\big)f'(t)$ **Chain Rule**

Example 11.2.7: Using derivative properties of vector-valued functions

Let $\vec{r}(t)=\langle t,t^2-1\rangle$ and let $\vec{u}(t)$ be the unit vector that points in the direction of $\vec{r}(t)$.

1. Graph $\vec{r}(t)$ and $\vec{u}(t)$ on the same axes, on $[-2,2]$.
2. Find $\vec{u}\,'(t)$ and sketch $\vec{u}\,'(-2)$, $\vec{u}\,'(-1)$ and $\vec{u}\,'(0)$. Sketch each with initial point the corresponding point on the graph of $\vec{u}$.

Solution

1. To form the unit vector that points in the direction of $\vec{r}$, we need to divide $\vec{r}(t)$ by its magnitude.

$$\|\vec{r}(t)\|=\sqrt{t^2+(t^2-1)^2}\quad\Rightarrow\quad\vec{u}(t)=\frac{1}{\sqrt{t^2+(t^2-1)^2}}\langle t,t^2-1\rangle. \tag{11.2.12}$$

   $\vec{r}(t)$ and $\vec{u}(t)$ are graphed in Figure 11.13. Note how the graph of $\vec{u}(t)$ forms part of a circle; this must be the case, as the length of $\vec{u}(t)$ is 1 for all t.

2. To compute $\vec{\mathbf{u}}\,'(t)$, we use Theorem 92, writing

$$\vec{\mathbf{u}}(t)=f(t)\vec{\mathbf{r}}(t),\quad \text{where}\quad f(t)=\frac{1}{\sqrt{t^2+(t^2-1)^2}}=\left(t^2+(t^2-1)^2\right)^{-1/2}. \tag{11.2.13}$$

(We *could* write

$$\vec{\mathbf{u}}(t)=\langle\frac{t}{\sqrt{t^2+(t^2-1)^2}},\frac{t^2-1}{\sqrt{t^2+(t^2-1)^2}}\rangle \tag{11.2.14}$$

and then take the derivative. It is a matter of preference; this latter method requires two applications of the Quotient Rule where our method uses the Product and Chain Rules.)

We find $f'(t)$ using the Chain Rule:

$$\begin{aligned}f'(t)&=-\frac{1}{2}\left(t^2+(t^2-1)^2\right)^{-3/2}\left(2t+2(t^2-1)(2t)\right)\\&=-\frac{2t(2t^2-1)}{2\left(\sqrt{t^2+(t^2-1)^2}\right)^3}\end{aligned}$$

We now find $\vec{\mathbf{u}}\,'(t)$ using part 3 of Theorem 92:

$$\begin{aligned}\vec{\mathbf{u}}\,'(t)&=f'(t)\vec{\mathbf{u}}(t)+f(t)\vec{\mathbf{u}}\,'(t)\\&=-\frac{2t(2t^2-1)}{2\left(\sqrt{t^2+(t^2-1)^2}\right)^3}\langle t,t^2-1\rangle+\frac{1}{\sqrt{t^2+(t^2-1)^2}}\langle 1,2t\rangle.\end{aligned}$$

This is admittedly very "messy;" such is usually the case when we deal with unit vectors. We can use this formula to compute $\vec{\mathbf{u}}\,(-2)$, $\vec{\mathbf{u}}\,(-1)$ and $\vec{\mathbf{u}}\,(0)$:

$$\begin{aligned}\vec{\mathbf{u}}\,(-2)&=\langle-\frac{15}{13\sqrt{13}},-\frac{10}{13\sqrt{13}}\rangle\approx\langle-0.320,-0.213\rangle\\\vec{\mathbf{u}}\,(-1)&=\langle 0,-2\rangle\\\vec{\mathbf{u}}\,(0)&=\langle 1,0\rangle\end{aligned}$$

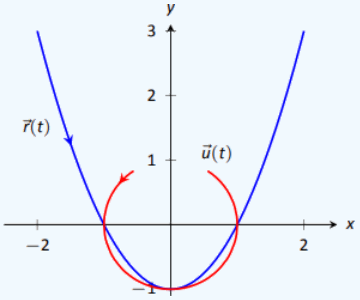


Figure 11.2.6: Graphing $\vec{\mathbf{r}}(t)$ and $\vec{\mathbf{u}}(t)$ in Example 11.2.7

Each of these is sketched in Figure 11.2.6. Note how the length of the vector gives an indication of how quickly the circle is being traced at that point. When $t=-2$, the circle is being drawn relatively slow; when $t=-1$, the circle is being traced much more quickly.

It is a basic geometric fact that a line tangent to a circle at a point P is perpendicular to the line passing through the center of the circle and P. This is illustrated in Figure 11.14; each tangent vector is perpendicular to the line that passes through its initial point and the center of the circle. Since the center of the circle is the origin, we can state this another way: $\vec{\mathbf{u}}\,'(t)$ is orthogonal to $\vec{\mathbf{u}}(t)$.

Recall that the dot product serves as a test for orthogonality: if $\vec{\mathbf{u}} \cdot \vec{\mathbf{v}} = 0$, then $\vec{\mathbf{u}}$ is orthogonal to $\vec{\mathbf{v}}$. Thus in the above example, $\vec{\mathbf{u}}(t) \cdot \vec{\mathbf{u}}\,'(t) = 0$.

This is true of any vector-valued function that has a constant length, that is, that traces out part of a circle. It has important implications later on, so we state it as a theorem (and leave its formal proof as an Exercise.)

THEOREM 93 Vector-Valued Functions of Constant Length

Let $\vec{\mathbf{r}}(t)$ be a differentiable vector-valued function on an open interval I of constant length. That is, $\|\vec{\mathbf{r}}(t)\| = c$ for all t in I (equivalently, $\vec{\mathbf{r}}(t) \cdot \vec{\mathbf{r}}(t) = c^2$ for all t in I). Then $\vec{\mathbf{r}}(t) \cdot \vec{\mathbf{r}}\,'(t) = 0$ for all t in I.

Integration

Indefinite and definite integrals of vector-valued functions are also evaluated component-wise.

THEOREM 94: Indefinite and Definite Integrals of Vector-Valued Functions

Let $\vec{\mathbf{r}}(t) = \langle f(t), g(t)\rangle$ be a vector-valued function in $\mathbb{R}^2$.

1. $\int \vec{\mathbf{r}}(t)\ dt = \langle \int f(t)\ dt, \int g(t)\ dt\rangle$
2. $\int_a^b \vec{\mathbf{r}}(t)\ dt = \langle \int_a^b f(t)\ dt, \int_a^b g(t)\ dt\rangle$

A similar statement holds for vector-valued functions in $\mathbb{R}^3$.

Example 11.2.8: Evaluating a definite integral of a vector-valued function

Let $\vec{\mathbf{r}}(t) = \langle e^{2t}, \sin t\rangle$. Evaluate $\int_0^1 \vec{\mathbf{r}}(t)\ dt$.

Solution

We follow Theorem 94.

$$\begin{aligned}\int_0^1 \vec{\mathbf{r}}(t)\ dt &= \int_0^1 \langle e^{2t}, \sin t\rangle\ dt \\ &= \langle \int_0^1 e^{2t}\ dt\ , \int_0^1 \sin t\ dt\rangle \\ &= \langle \frac{1}{2}e^{2t}\Big|_0^1\ , -\cos t\Big|_0^1\rangle \\ &= \langle \frac{1}{2}(e^2-1)\ , -\cos(1)+1\rangle \\ &\approx \langle 3.19, 0.460\rangle.\end{aligned}$$

Example 11.2.9: Solving an initial value problem

Let $\vec{\mathbf{r}}\,''(t) = \langle 2, \cos t, 12t\rangle$. Find $\vec{\mathbf{r}}(t)$ where:

1. $\vec{\mathbf{r}}(0) = \langle -7, -1, 2\rangle$ and
2. $\vec{\mathbf{r}}\,'(0) = \langle 5, 3, 0\rangle$.

Solution

Knowing $\vec{\mathbf{r}}\,''(t) = \langle 2, \cos t, 12t\rangle$, we find $\vec{\mathbf{r}}\,'(t)$ by evaluating the indefinite integral.

$$\begin{aligned}\int \vec{\mathbf{r}}\,''(t)dt &= \langle \int 2dt\ , \int \cos t dt\ , \int 12tdt\rangle \\ &= \langle 2t + C_1, \sin t + C_2, 6t^2 + C_3\rangle \\ &= \langle 2t, \sin t, 6t^2\rangle + \langle C_1, C_2, C_3\rangle \\ &= \langle 2t, \sin t, 6t^2\rangle + \vec{\mathbf{C}}_1.\end{aligned}$$

Note how each indefinite integral creates its own constant which we collect as one constant vector $\vec{\mathbf{C}}_1$. Knowing $\vec{\mathbf{r}}\,'(0) = \langle 5, 3, 0\rangle$ allows us to solve for $\vec{\mathbf{C}}_1$:

$$\begin{aligned}\vec{\mathbf{r}}'(t) &= \langle 2t, \sin t, 6t^2\rangle + \vec{\mathbf{C}}_1\\ \vec{\mathbf{r}}'(0) &= \langle 0,0,0\rangle + \vec{\mathbf{C}}_1\\ \langle 5,3,0\rangle &= \vec{\mathbf{C}}_1.\end{aligned}$$

So $\vec{\mathbf{r}}'(t) = \langle 2t, \sin t, 6t^2\rangle + \langle 5,3,0\rangle = \langle 2t+5, \sin t + 3, 6t^2\rangle$. To find $\vec{\mathbf{r}}(t)$, we integrate once more.

$$\begin{aligned}\int \vec{\mathbf{r}}'(t)\ dt &= \langle \int 2t+5\ dt, \int \sin t + 3\ dt, \int 6t^2 dt\rangle\\ &= \langle t^2+5t, -\cos t + 3t, 2t^3\rangle + \vec{\mathbf{C}}_2.\end{aligned}$$

With $\vec{\mathbf{r}}(0) = \langle -7,-1,2\rangle$, we solve for $\vec{\mathbf{C}}_2$:

$$\begin{aligned}\vec{\mathbf{r}}(t) &= \langle t^2+5t, -\cos t + 3t, 2t^3\rangle + \vec{\mathbf{C}}_2\\ \vec{\mathbf{r}}(0) &= \langle 0,-1,0\rangle + \vec{\mathbf{C}}_2\\ \langle -7,-1,2\rangle &= \langle 0,-1,0\rangle + \vec{\mathbf{C}}_2\\ \langle -7,0,2\rangle &= \vec{\mathbf{C}}_2.\end{aligned}$$

So $\vec{\mathbf{r}}(t) = \langle t^2+5t, -\cos t + 3t, 2t^3\rangle + \langle -7,0,2\rangle = \langle t^2+5t-7, -\cos t + 3t, 2t^3+2\rangle$.

What does the integration of a vector-valued function *mean*? There are many applications, but none as direct as "the area under the curve" that we used in understanding the integral of a real-valued function.

A key understanding for us comes from considering the integral of a derivative:

$$\int_a^b \vec{\mathbf{r}}'(t)dt = \vec{\mathbf{r}}(t)\Big|_a^b = \vec{\mathbf{r}}(b) - \vec{\mathbf{r}}(a). \tag{11.2.15}$$

Integrating a *rate of change* function gives *displacement*.

Noting that vector-valued functions are closely related to parametric equations, we can describe the arc length of the graph of a vector-valued function as an integral. Given parametric equations $x = f(t)$, $y = g(t)$, the arc length on $[a,b]$ of the graph is

$$\text{Arc Length} = \int_a^b \sqrt{f'(t)^2 + g'(t)^2}dt, \tag{11.2.16}$$

as stated in Theorem 82 in Section 9.3. If $\vec{\mathbf{r}}(t) = \langle f(t), g(t)\rangle$, note that $\sqrt{f'(t)^2 + g'(t)^2} = \|\vec{\mathbf{r}}'(t)\|$. Therefore we can express the arc length of the graph of a vector-valued function as an integral of the magnitude of its derivative.

THEOREM 95: Arc Length of a Vector-Valued Function

Let $\vec{\mathbf{r}}(t)$ be a vector-valued function where $\vec{\mathbf{r}}'(t)$ is continuous on $[a,b]$. The arc length L of the graph of $\vec{\mathbf{r}}(t)$ is

$$L = \int_a^b \|\vec{\mathbf{r}}'(t)\|\ dt. \tag{11.2.17}$$

Note that we are actually integrating a scalar-function here, not a vector-valued function.

The next section takes what we have established thus far and applies it to objects in motion. We will let $\vec{\mathbf{r}}(t)$ describe the path of an object in the plane or in space and will discover the information provided by $\vec{\mathbf{r}}'(t)$ and $\vec{\mathbf{r}}''(t)$.

This page titled 11.2: Calculus and Vector-Valued Functions is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

11.3: The Calculus of Motion

A common use of vector--valued functions is to describe the motion of an object in the plane or in space. A **position function** $\vec{\textbf{r}}(t)$ gives the position of an object at **time** t. This section explores how derivatives and integrals are used to study the motion described by such a function.

Definition 73: Velocity, Speed and Acceleration

Let $\vec{\textbf{r}}(t)$ be a position function in $\mathbb{R}^2$ or $\mathbb{R}^3$.

- **Velocity**, denoted $\vec{\textbf{v}}(t)$, is the instantaneous rate of position change; that is,

$$\vec{\textbf{v}}(t)=\vec{\textbf{r}}\,'(t). \tag{11.3.1}$$

- **Speed** is the magnitude of velocity,

$$\text{speed}=\|\vec{\textbf{v}}(t)\|. \tag{11.3.2}$$

- **Acceleration**, denoted $\vec{\textbf{a}}(t)$, is the instantaneous rate of velocity change; that is,

$$\vec{\textbf{a}}(t)=\vec{\textbf{v}}\,'(t)=\vec{\textbf{r}}\,''(t). \tag{11.3.3}$$

Example 11.3.1: Finding velocity and acceleration

An object is moving with position function $\vec{\textbf{r}}(t)=\langle t^2-t, t^2+t\rangle$, $-3\le t\le 3$, where distances are measured in feet and time is measured in seconds.

1. Find $\vec{\textbf{v}}(t)$ and $\vec{\textbf{a}}(t)$.
2. Sketch $\vec{\textbf{r}}(t)$; plot $\vec{\textbf{v}}(-1)$, $\vec{\textbf{a}}(-1)$, $\vec{\textbf{v}}(1)$ and $\vec{\textbf{a}}(1)$, each with their initial point at their corresponding point on the graph of $\vec{\textbf{r}}(t)$.
3. When is the object's speed minimized?

Solution

1. Taking derivatives, we find

$$\vec{\textbf{v}}(t)=\vec{\textbf{r}}\,'(t)=\langle 2t-1, 2t+1\rangle$$

and

$$\vec{\textbf{a}}(t)=\vec{\textbf{r}}\,''(t)=\langle 2,2\rangle.$$

Note that acceleration is constant.

2. $\vec{\textbf{v}}(-1)=\langle -3,-1\rangle$, $\vec{\textbf{a}}(-1)=\langle 2,2\rangle$; $\vec{\textbf{v}}(1)=\langle 1,3\rangle$, $\vec{\textbf{a}}(1)=\langle 2,2\rangle$. These are plotted with $\vec{\textbf{r}}(t)$ in Figure $11.3.1a$.
We can think of acceleration as "pulling" the velocity vector in a certain direction. At $t=-1$, the velocity vector points down and to the left; at $t=1$, the velocity vector has been pulled in the $\langle 2,2\rangle$ direction and is now pointing up and to the right. In Figure 11.15(b) we plot more velocity/acceleration vectors, making more clear the effect acceleration has on velocity.
Since $\vec{\textbf{a}}(t)$ is constant in this example, as t grows large $\vec{\textbf{v}}(t)$ becomes almost parallel to $\vec{\textbf{a}}(t)$. For instance, when $t=10$, $\vec{\textbf{v}}(10)$ $=\langle 19,21\rangle$, which is nearly parallel to $\langle 2,2\rangle$.
3. The object's speed is given by

$$\|\vec{\textbf{v}}(t)\|=\sqrt{(2t-1)^2+(2t+1)^2}=\sqrt{8t^2+2}. \tag{11.3.4}$$

To find the minimal speed, we could apply calculus techniques (such as set the derivative equal to 0 and solve for t, etc.) but we can find it by inspection. Inside the square root we have a quadratic which is minimized when $t=0$. Thus the speed is minimized at $t=0$, with a speed of $\sqrt{2}$ ft/s.

The graph in Figure $11.3.1b$ also implies speed is minimized here. The filled dots on the graph are located at integer values of t between -3 and 3. Dots that are far apart imply the object traveled a far distance in 1 second, indicating high speed; dots that are close together imply the object did not travel far in 1 second, indicating a low speed. The dots are closest together near $t=0$, implying the speed is minimized near that value.

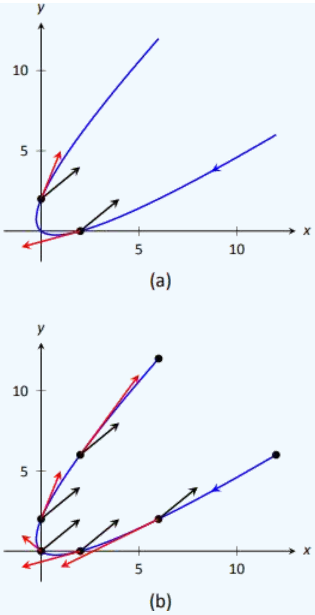


Figure 11.3.1: Graphing the position, velocity and acceleration of an object in Example 11.3.1

Example 11.3.2: Analyzing Motion

Two objects follow an identical path at different rates on $[-1,1]$. The position function for Object 1 is $\vec{\mathbf{r}}_1(t) = \langle t, t^2\rangle$; the position function for Object 2 is $\vec{\mathbf{r}}_2(t) = \langle t^3, t^6\rangle$, where distances are measured in feet and time is measured in seconds. Compare the velocity, speed and acceleration of the two objects on the path.

Solution

We begin by computing the velocity and acceleration function for each object:

$$\begin{aligned} \vec{\mathbf{v}}_1(t) &= \langle 1, 2t\rangle & \vec{\mathbf{v}}_2(t) &= \langle 3t^2, 6t^5\rangle \\ \vec{\mathbf{a}}_1(t) &= \langle 0, 2\rangle & \vec{\mathbf{a}}_2(t) &= \langle 6t, 30t^4\rangle \end{aligned}$$

We immediately see that Object 1 has constant acceleration, whereas Object 2 does not.

At $t = -1$, we have $\vec{\mathbf{v}}_1(-1) = \langle 1, -2\rangle$ and $\vec{\mathbf{v}}_2(-1) = \langle 3, -6\rangle$; the velocity of Object 2 is three times that of Object 1 and so it follows that the speed of Object 2 is three times that of Object 1 ($3\sqrt{5}$ ft/s compared to $\sqrt{5}$ ft/s.)

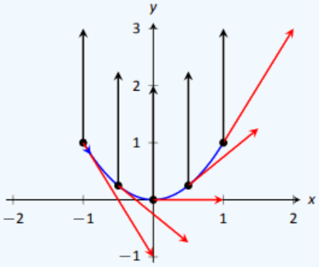
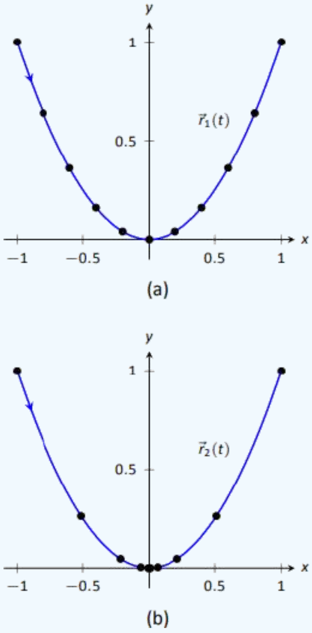


Figure 11.3.2: Plotting velocity and acceleration vectors for Object 1 in Example 11.3.2

At $t = 0$, the velocity of Object 1 is $\vec{\mathbf{v}}(1) = \langle 1, 0\rangle$ and the velocity of Object 2 is $\vec{0}$! This tells us that Object 2 comes to a complete stop at $t = 0$.

In Figure 11.3.2, we see the velocity and acceleration vectors for Object 1 plotted for $t = -1, -1/2, 0, 1/2$ and $t = 1$. Note again how the constant acceleration vector seems to "pull" the velocity vector from pointing down, right to up, right. We could plot the analogous picture for Object 2, but the velocity and acceleration vectors are rather large ($\vec{\mathbf{a}}_2(-1) = \langle -6, 30\rangle$!)

Instead, we simply plot the locations of Object 1 and 2 on intervals of $1/10^{\text{th}}$ of a second, shown in Figure $11.3.3a$ and $11.3.3b$. Note how the x-values of Object 1 increase at a steady rate. This is because the x-component of $\vec{\mathbf{a}}(t)$ is 0; there is no acceleration in the x-component. The dots are not evenly spaced; the object is moving faster near $t = -1$ and $t = 1$ than near $t = 0$.



(a)

(b)

Figure $11.3.3$: Comparing the positions of Objects 1 and 2 in Example 11.3.2

In Figure $11.3.3b$, we see the points plotted for Object 2. Note the large change in position from $t = -1$ to $t = -0.9$; the object starts moving very quickly. However, it slows considerably at it approaches the origin, and comes to a complete stop at $t = 0$. While it looks like there are 3 points near the origin, there are in reality 5 points there.

Since the objects begin and end at the same location, the have the same displacement. Since they begin and end at the same time, with the same displacement, they have they have the same average rate of change (i.e, they have the same average velocity). Since they follow the same path, they have the same distance traveled. Even though these three measurements are the same, the objects obviously travel the path in very different ways.

Example $11.3.3$: Analyzing the motion of a whirling ball on a string

A young boy whirls a ball, attached to a string, above his head in a counter-clockwise circle. The ball follows a circular path and makes 2 revolutions per second. The string has length 2ft.

1. Find the position function $\vec{\mathbf{r}}(t)$ that describes this situation.
2. Find the acceleration of the ball and derive a physical interpretation of it.
3. A tree stands 10 ft in front of the boy. At what t-values should the boy release the string so that the ball hits the tree?

Solution

1. The ball whirls in a circle. Since the string is 2 ft long, the radius of the circle is 2. The position function $\vec{\mathbf{r}}(t) = \langle 2\cos t, 2\sin t\rangle$ describes a circle with radius 2, centered at the origin, but makes a full revolution every 2π seconds, not two revolutions per second. We modify the period of the trigonometric functions to be 1/2 by multiplying t by 4π. The final position function is thus
$$\vec{\mathbf{r}}(t) = \langle 2\cos(4\pi t), 2\sin(4\pi t)\rangle. \tag{11.3.5}$$
(Plot this for $0 \le t \le 1/2$ to verify that one revolution is made in 1/2 a second.)
2. To find $\vec{\mathbf{a}}(t)$, we derive $\vec{\mathbf{r}}(t)$ twice.
$$\begin{aligned}\vec{\mathbf{v}}(t) = \vec{\mathbf{r}}'(t) &= \langle -8\pi\sin(4\pi t), 8\pi\cos(4\pi t)\rangle \\ \vec{\mathbf{a}}(t) = \vec{\mathbf{r}}''(t) &= \langle -32\pi^2\cos(4\pi t), -32\pi^2\sin(4\pi t)\rangle \\ &= -32\pi^2\langle\cos(4\pi t), \sin(4\pi t)\rangle.\end{aligned}$$

Note how $\vec{\mathbf{a}}(t)$ is parallel to $\vec{\mathbf{r}}(t)$, but has a different magnitude and points in the opposite direction. Why is this?

Recall the classic physics equation, "Force $=$ mass $\times$ acceleration." A force acting on a mass induces acceleration (i.e., the mass moves); acceleration acting on a mass induces a force (gravity gives our mass a *weight*). Thus force and acceleration are closely related. A moving ball "wants" to travel in a straight line. Why does the ball in our example move in a circle? It is attached to the boy's hand by a string. The string applies a force to the ball, affecting it's motion: the string *accelerates* the ball. This is not acceleration in the sense of "it travels faster;" rather, this acceleration is changing the velocity of the ball. In what direction is this force/acceleration being applied? In the direction of the string, towards the boy's hand.

The magnitude of the acceleration is related to the speed at which the ball is traveling. A ball whirling quickly is rapidly changing direction/velocity. When velocity is changing rapidly, the acceleration must be "large."

3. When the boy releases the string, the string no longer applies a force to the ball, meaning acceleration is $\vec{0}$ and the ball can now move in a straight line in the direction of $\vec{\mathbf{v}}(t)$.

   Let $t=t_0$ be the time when the boy lets go of the string. The ball will be at $\vec{\mathbf{r}}(t_0)$, traveling in the direction of $\vec{\mathbf{v}}(t_0)$. We want to find t_0 so that this line contains the point $(0,10)$ (since the tree is 10ft directly in front of the boy).

   There are many ways to find this time value. We choose one that is relatively simple computationally. As shown in Figure $11.3.4$, the vector from the release point to the tree is $\langle 0,10\rangle-\vec{\mathbf{r}}(t_0)$. This line segment is tangent to the circle, which means it is also perpendicular to $\vec{\mathbf{r}}(t_0)$ itself, so their dot product is 0.

$$\begin{aligned}
\vec{\mathbf{r}}(t_0)\cdot\big(\langle 0,10\rangle-\vec{\mathbf{r}}(t_0)\big)&=0\\
\langle 2\cos(4\pi t_0),2\sin(4\pi t_0)\rangle\cdot\langle -2\cos(4\pi t_0),10-2\sin(4\pi t_0)\rangle&=0\\
-4\cos^2(4\pi t_0)+20\sin(4\pi t_0)-4\sin^2(4\pi t_0)&=0\\
20\sin(4\pi t_0)-4&=0\\
\sin(4\pi t_0)&=1/5\\
4\pi t_0&=\sin^{-1}(1/5)\\
4\pi t_0&\approx 0.2+2\pi n,
\end{aligned}$$

   where n is an integer. Solving for t_0 we have:

$$t_0\approx 0.016+n/2$$

   This is a wonderful formula. Every 1/2 second after $t=0.016$s the boy can release the string (since the ball makes 2 revolutions per second, he has two chances each second to release the ball).

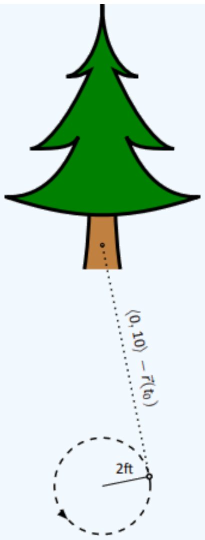


Figure $11.3.4$: Modeling the flight of a ball in Example 11.3.3

Example $11.3.4$: Analyzing motion in space

An object moves in a spiral with position function $\vec{\mathbf{r}}(t) = \langle \cos t, \sin t, t\rangle$, where distances are measured in meters and time is in minutes. Describe the object's speed and acceleration at time t.

Solution

With $\vec{\mathbf{r}}(t) = \langle \cos t, \sin t, t\rangle$, we have:

$$\begin{aligned} \vec{\mathbf{v}}(t) &= \langle -\sin t, \cos t, 1\rangle \quad \text{and} \\ \vec{\mathbf{a}}(t) &= \langle -\cos t, -\sin t, 0\rangle. \end{aligned}$$

The speed of the object is $\|\vec{\mathbf{v}}(t)\| = \sqrt{(-\sin t)^2 + \cos^2 t + 1} = \sqrt{2}$m/min; it moves at a constant speed. Note that the object does not accelerate in the z-direction, but rather moves up at a constant rate of 1m/min.

The objects in Examples 11.3.3 and 11.3.4 traveled at a constant speed. That is, $\|\vec{\mathbf{v}}(t)\| = c$ for some constant c. Recall Theorem 93, which states that if a vector--valued function $\vec{\mathbf{r}}(t)\backslash$ has constant length, then $\vec{\mathbf{r}}(t)\backslash$ is perpendicular to its derivative: $\vec{\mathbf{r}}(t) \cdot \vec{\mathbf{r}}'(t) = 0$. In these examples, the velocity function has constant length, therefore we can conclude that the velocity is perpendicular to the acceleration: $\vec{\mathbf{v}}(t) cdot \vec{\mathbf{a}}(t) = 0$. A quick check verifies this.

There is an intuitive understanding of this. If acceleration is parallel to velocity, then it is only affecting the object's speed; it does not change the direction of travel. (For example, consider a dropped stone. Acceleration and velocity are parallel -- straight down -- and the direction of velocity never changes, though speed does increase.) If acceleration is not perpendicular to velocity, then there is some acceleration in the direction of travel, influencing the speed. If speed is constant, then acceleration must be orthogonal to velocity, as it then only affects direction, and not speed.

key idea: 52 Objects With Constant Speed

If an object moves with constant speed, then its velocity and acceleration vectors are orthogonal. That is, $\vec{\mathbf{v}}(t) cdot \vec{\mathbf{a}}(t) = 0$.

Projectile Motion

An important application of vector--valued position functions is *projectile motion*: the motion of objects under only the influence of gravity. We will measure time in seconds, and distances will either be in meters or feet. We will show that we can completely describe the path of such an object knowing its initial position and initial velocity (i.e., where it *is* and where it *is going*.)

Suppose an object has initial position $\vec{\mathbf{r}}(0) = \langle x_0, y_0\rangle$ and initial velocity $\vec{\mathbf{v}}(0) = \langle v_x, v_y\rangle$. It is customary to rewrite $\vec{\mathbf{v}}(0)$ in terms of its speed v_0 and direction $\vec{\mathbf{u}}$, where $\vec{\mathbf{u}}$ is a unit vector. Recall all unit vectors in $\mathbb{R}^2$ can be written as $\langle \cos\theta, \sin\theta\rangle$, where θ is an angle measure counter--clockwise from the x-axis. (We refer to θ as the **angle of elevation**.) Thus $\vec{\mathbf{v}}(0) = v_0\langle \cos\theta, \sin\theta\rangle$.

Since the acceleration of the object is known, namely $\vec{\mathbf{a}}(t) = \langle 0, -g\rangle$, where g is the gravitational constant, we can find $\vec{\mathbf{r}}(t)$ knowing our two initial conditions. We first find \(\vecs v (t)):

Note

In this text we use $g = 32\text{ft/s}$ when using Imperial units, and $g = 9.8\text{m/s}$ when using SI units.

$$\begin{aligned}\vec{\mathbf{v}}(t) &= \int \vec{\mathbf{a}}(t)dt\\ \vec{\mathbf{v}}(t) &= \int \langle 0, -g\rangle dt\\ \vec{\mathbf{v}}(t) &= \langle 0, -gt\rangle + \vec{\mathbf{C}}.\end{aligned}$$

Knowing $\vec{\mathbf{v}}(0) = v_0\langle \cos\theta, \sin\theta\rangle$, we have $\vec{\mathbf{C}} = v_0\langle \cos t, \sin t\rangle$ and so

$$\vec{\mathbf{v}}(t) = \langle v_0\cos\theta, -gt + v_0\sin\theta\rangle. \tag{11.3.6}$$

We integrate once more to find $\vec{\mathbf{r}}(t)$:

$$\begin{aligned}\vec{\mathbf{r}}(t) &= \int \vec{\mathbf{v}}(t)dt\\ \vec{\mathbf{r}}(t) &= \int \langle v_0\cos\theta, -gt + v_0\sin\theta\rangle dt\\ \vec{\mathbf{r}}(t) &= \langle \big(v_0\cos\theta\big)t, -\frac{1}{2}gt^2 + \big(v_0\sin\theta\big)t\rangle + \vec{\mathbf{C}}.\end{aligned}$$

Knowing $\vec{\mathbf{r}}(0) = \langle x_0, y_0\rangle$, we conclude $\vec{\mathbf{C}} = \langle x_0, y_0\rangle$ and

$$\vec{\mathbf{r}}(t) = \langle \big(v_0\cos\theta\big)t + x_0\ , -\frac{1}{2}gt^2 + \big(v_0\sin\theta\big)t + y_0\ \rangle.$$

key idea 53: Projectile Motion

The position function of a projectile propelled from an initial position of $\vec{\mathbf{r}}_0 = \langle x_0, y_0\rangle$, with initial speed v_0, with angle of elevation θ and neglecting all accelerations but gravity is

$$\vec{\mathbf{r}}(t) = \langle \big(v_0\cos\theta\big)t + x_0\ , -\frac{1}{2}gt^2 + \big(v_0\sin\theta\big)t + y_0\ \rangle. \tag{11.3.7}$$

Letting $\vec{\mathbf{v}}_0 = v_0\langle \cos\theta, \sin\theta\rangle$, (\vecs r (t)\) can be written as

$$\vec{\mathbf{r}}(t) = \langle 0, -\frac{1}{2}gt^2\rangle + \vec{\mathbf{v}}_0 t + \vec{\mathbf{r}}_0. \tag{11.3.8}$$

We demonstrate how to use this position function in the next two examples.

Example 11.3.5: Projectile Motion

Sydney shoots her Red Ryder bb gun across level ground from an elevation of 4ft, where the barrel of the gun makes a 5° angle with the horizontal. Find how far the bb travels before landing, assuming the bb is fired at the advertised rate of 350ft/s and ignoring air resistance.

Solution

A direct application of Key Idea 53 gives

$$\begin{aligned}\vec{\mathbf{r}}(t) &= \langle (350\cos 5^\circ)t, -16t^2 + (350\sin 5^\circ)t + 4\rangle\\ &\approx \langle 346.67t, -16t^2 + 30.50t + 4\rangle,\end{aligned}$$

where we set her initial position to be $\langle 0, 4\rangle$.
We need to find *when* the bb lands, then we can find *where*. We accomplish this by setting the y-component equal to 0 and solving for t:

$$\begin{aligned} -16t^2+30.50t+4&=0\\ t&=\frac{-30.50\pm\sqrt{30.50^2-4(-16)(4)}}{-32}\\ t&\approx 2.03s. \end{aligned}$$

(We discarded a negative solution that resulted from our quadratic equation.)

We have found that the bb lands 2.03s after firing; with $t=2.03$, we find the x-component of our position function is $346.67(2.03)=703.74$ft. The bb lands about 704 feet away.

Example 11.3.6: Projectile Motion

Alex holds his sister's bb gun at a height of 3ft and wants to shoot a target that is 6ft above the ground, 25ft away. At what angle should he hold the gun to hit his target? (We still assume the muzzle velocity is 350ft/s.)

Solution

The position function for the path of Alex's bb is

$$\vec{\mathbf{r}}(t)=\langle(350\cos\theta)t,-16t^2+(350\sin\theta)t+3\rangle.$$

We need to find θ so that $\vec{\mathbf{r}}(t)$ =\langle 25,6\rangle\) for some value of t. That is, we want to find θ and t such that

$$(350\cos\theta)t=25\quad\text{and}\quad -16t^2+(350\sin\theta)t+3=6.$$

This is not trivial (though not "hard"). We start by solving each equation for $\cos\theta$ and $\sin\theta$, respectively.

$$\cos\theta=\frac{25}{350t}\quad\text{and}\quad\sin\theta=\frac{3+16t^2}{350t}.$$

Using the Pythagorean Identity $\cos^2\theta+\sin^2\theta=1$, we have

$$\left(\frac{25}{350t}\right)^2+\left(\frac{3+16t^2}{350t}\right)^2=1$$

Multiply both sides by $(350t)^2$:

$$\begin{aligned} 25^2+(3+16t^2)^2&=350^2t^2\\ 256t^4-122,404t^2+634&=0. \end{aligned}$$

This is a quadratic *in* t^2. That is, we can apply the quadratic formula to find t^2, then solve for t itself.

$$\begin{aligned} t^2&=\frac{122,404\pm\sqrt{122,404^2-4(256)(634)}}{512}\\ t^2&=0.0052,\ \ 478.135\\ t&=\pm0.072,\ \ \pm21.866 \end{aligned}$$

Clearly the negative t values do not fit our context, so we have $t=0.072$ and $t=21.866$. Using $\cos\theta=25/(350t)$, we can solve for θ:

$$\begin{aligned} \theta&=\cos^{-1}\left(\frac{25}{350\cdot0.072}\right)\quad\text{and}\quad\cos^{-1}\left(\frac{25}{350\cdot21.866}\right)\\ \theta&=7.03^\circ\quad\text{and}\quad 89.8^\circ. \end{aligned}$$

Alex has two choices of angle. He can hold the rifle at an angle of about 7° with the horizontal and hit his target 0.07s after firing, or he can hold his rifle almost straight up, with an angle of 89.8°, where he'll hit his target about 22s later. The first option is clearly the option he should choose.

Distance Traveled

Consider a driver who sets her cruise--control to 60mph, and travels at this speed for an hour. We can ask:

1. How far did the driver travel?
2. How far from her starting position is the driver?

The first is easy to answer: she traveled 60 miles. The second is impossible to answer with the given information. We do not know if she traveled in a straight line, on an oval racetrack, or along a slowly--winding highway.

This highlights an important fact: to compute distance traveled, we need only to know the speed, given by $\|\vec{\mathbf{v}}(t)\|$.

theorem 96: Distance Traveled

Let $\vec{\mathbf{v}}(t)$ be a velocity function for a moving object. The distance traveled by the object on $[a, b]$ is:

$$\text{distance traveled} = \int_a^b \|\vec{\mathbf{v}}(t)\| dt. \tag{11.3.9}$$

Note that this is just a restatement of Theorem 95: arc length is the same as distance traveled, just viewed in a different context.

Example $11.3.7$: Distance Traveled, Displacement, and Average Speed

A particle moves in space with position function $\vec{\mathbf{r}}(t) = \langle t, t^2, \sin(\pi t)\rangle$ on $[-2, 2]$, where t is measured in seconds and distances are in meters. Find:

1. The distance traveled by the particle on $[-2, 2]$.
2. The displacement of the particle on $[-2, 2]$.
3. The particle's average speed.

Solution

1. We use Theorem 96 to establish the integral:

$$\begin{aligned}\text{distance traveled} &= \int_{-2}^{2} \|\vec{\mathbf{v}}(t)\| dt \\ &= \int_{-2}^{2} \sqrt{1 + (2t)^2 + \pi^2 \cos^2(\pi t)} dt.\end{aligned}$$

   This cannot be solved in terms of elementary functions so we turn to numerical integration, finding the distance to be 12.88m.

2. The displacement is the vector

$$\vec{\mathbf{r}}(2) - \vec{\mathbf{r}}(-2) = \langle 2, 4, 0\rangle - \langle -2, 4, 0\rangle = \langle 4, 0, 0\rangle.$$

   That is, the particle ends with an x-value increased by 4 and with y- and z-values the same (see Figure $11.3.5$).

3. We found above that the particle traveled 12.88m over 4 seconds. We can compute average speed by dividing: 12.88/4 = 3.22m/s.

   We should also consider Definition 22 of Section 5.4, which says that the average value of a function f on $[a, b]$ is $\frac{1}{b-a} \int_a^b f(x)\ dx$. In our context, the average value of the speed is

$$\text{average speed} = \frac{1}{2-(-2)} \int_{-2}^{2} \|\vec{\mathbf{v}}(t)\| dt \approx \frac{1}{4} 12.88 = 3.22 \text{m/s}.$$

   Note how the physical context of a particle traveling gives meaning to a more abstract concept learned earlier.

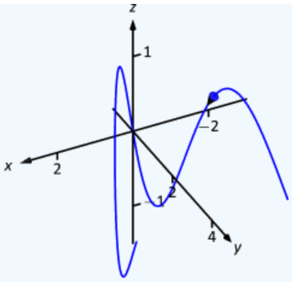


Figure 11.3.5: The path of the particle in Example 11.3.7

In Definition 22 of Chapter 5 we defined the average value of a function $f(x)$ on $[a,b]$ to be

$$\frac{1}{b-a}\int_a^b f(x)dx. \tag{11.3.10}$$

Note how in Example 11.3.7 we computed the average speed as

$$\frac{\text{distance traveled}}{\text{travel time}} = \frac{1}{2-(-2)}\int_{-2}^{2} \|\vec{\mathbf{v}}(t)\| dt; \tag{11.3.11}$$

that is, we just found the average value of $\|\vec{\mathbf{v}}(t)\|$ on $[-2,2]$.

Likewise, given position function $\vec{\mathbf{r}}(t)$, the average velocity on $[a,b]$ is

$$\frac{\text{displacement}}{\text{travel time}} = \frac{1}{b-a}\int_a^b \vec{\mathbf{r}}\,'(t)dt = \frac{\vec{\mathbf{r}}(b)-\vec{\mathbf{r}}(a)}{b-a}; \tag{11.3.12}$$

that is, it is the average value of $\vec{\mathbf{r}}\,'(t)$, or $\vec{\mathbf{v}}(t)$, on $[a,b]$.

KEY IDEA 54: Average Speed, Average Velocity

Let $\vec{\mathbf{r}}(t)$ be a continuous position function on an open interval I containing $a<b$.

- The **average speed** is:

$$\frac{\text{distance traveled}}{\text{travel time}} = \frac{\int_a^b \|\vec{\mathbf{v}}(t)\| dt}{b-a} = \frac{1}{b-a}\int_a^b \|\vec{\mathbf{v}}(t)\| dt. \tag{11.3.13}$$

- The **average velocity** is:

$$\frac{\text{displacement}}{\text{travel time}} = \frac{\int_a^b \vec{\mathbf{r}}\,'(t)dt}{b-a} = \frac{1}{b-a}\int_a^b \vec{\mathbf{r}}\,'(t)dt. \tag{11.3.14}$$

The next two sections investigate more properties of the graphs of vector--valued functions and we'll apply these new ideas to what we just learned about motion.

This page titled 11.3: The Calculus of Motion is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

11.4: Unit Tangent and Normal Vectors

Unit Tangent Vector

Given a smooth vector-valued function $\vec{\mathbf{r}}(t)$, we defined in Definition 71 that any vector parallel to $\vec{\mathbf{r}}\,'(t_0)$ is *tangent* to the graph of $\vec{\mathbf{r}}(t)$ at $t=t_0$. It is often useful to consider just the *direction* of $\vec{\mathbf{r}}\,'(t)$ and not its magnitude. Therefore we are interested in the unit vector in the direction of $\vec{\mathbf{r}}\,'(t)$. This leads to a definition.

Definition 11.4.1: Unit Tangent Vector

Let $\vec{\mathbf{r}}(t)$ be a smooth function on an open interval I. The unit tangent vector $\vec{\mathbf{T}}(t)$ is

$$\vec{\mathbf{T}}(t)=\frac{1}{\|\vec{\mathbf{r}}\,'(t)\|}\vec{\mathbf{r}}\,'(t). \tag{11.4.1}$$

Example 11.4.1: Computing the unit tangent vector

Let $\vec{\mathbf{r}}(t)=\langle 3\cos t, 3\sin t, 4t\rangle$. Find $\vec{\mathbf{T}}(t)$ and compute $\vec{\mathbf{T}}(0)$ and $\vec{\mathbf{T}}(1)$.

Solution

We apply Definition 11.4.1 to find $\vec{\mathbf{T}}(t)$.

$$\begin{aligned}\vec{\mathbf{T}}(t)&=\frac{1}{\|\vec{\mathbf{r}}\,'(t)\|}\vec{\mathbf{r}}\,'(t)\\ &=\frac{1}{\sqrt{(-3\sin t)^2+(3\cos t)^2+4^2}}\langle -3\sin t, 3\cos t, 4\rangle\\ &=\langle -\frac{3}{5}\sin t, \frac{3}{5}\cos t, \frac{4}{5}\rangle.\end{aligned}$$

We can now easily compute $\vec{\mathbf{T}}(0)$ and $\vec{\mathbf{T}}(1)$:

$$\begin{aligned}\vec{\mathbf{T}}(0)&=\langle 0, \frac{3}{5}, \frac{4}{5}\rangle\\ \vec{\mathbf{T}}(1)&=\langle -\frac{3}{5}\sin 1, \frac{3}{5}\cos 1, \frac{4}{5}\rangle \approx \langle -0.505, 0.324, 0.8\rangle.\end{aligned}$$

These are plotted in Figure 11.4.1 with their initial points at $\vec{\mathbf{r}}(0)$ and $\vec{\mathbf{r}}(1)$, respectively. (They look rather "short" since they are only length 1.)

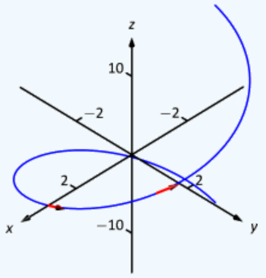


Figure 11.4.1: Plotting unit tangent vectors in Example 11.4.1.

The unit tangent vector $\vec{\mathbf{T}}(t)$ always has a magnitude of 1, though it is sometimes easy to doubt that is true. We can help solidify this thought in our minds by computing $\|\vec{\mathbf{T}}(1)\|$:

$$\|\vec{\mathbf{T}}(1)\| \approx \sqrt{(-0.505)^2+0.324^2+0.8^2}=1.000001.$$

We have rounded in our computation of $\vec{\mathbf{T}}(1)$, so we don't get 1 exactly. We leave it to the reader to use the exact representation of $\vec{\mathbf{T}}(1)$ to verify it has length 1.

In many ways, the previous example was "too nice." It turned out that $\vec{\mathbf{r}}\,'(t)$ was always of length 5. In the next example the length of $\vec{\mathbf{r}}\,'(t)$ is variable, leaving us with a formula that is not as clean.

Example $11.4.2$: Computing the unit tangent vector

Let $\vec{\mathbf{r}}(t) = \langle t^2 - t, t^2 + t\rangle$. Find $\vec{\mathbf{T}}(t)$ and compute $\vec{\mathbf{T}}(0)$ and $\vec{\mathbf{T}}(1)$.

Solution

We find $\vec{\mathbf{r}}\,'(t) = \langle 2t-1, 2t+1\rangle$, and

$$\|\vec{\mathbf{r}}\,'(t)\| = \sqrt{(2t-1)^2 + (2t+1)^2} = \sqrt{8t^2+2}. \tag{11.4.2}$$

Therefore

$$\vec{\mathbf{T}}(t) = \frac{1}{\sqrt{8t^2+2}}\langle 2t-1, 2t+1\rangle = \langle \frac{2t-1}{\sqrt{8t^2+2}}, \frac{2t+1}{\sqrt{8t^2+2}}\rangle. \tag{11.4.3}$$

When $t=0$, we have $\vec{\mathbf{T}}(0) = \langle -1/\sqrt{2}, 1/\sqrt{2}\rangle$; when $t=1$, we have $\vec{\mathbf{T}}(1) = \langle 1/\sqrt{10}, 3/\sqrt{10}\rangle$. We leave it to the reader to verify each of these is a unit vector. They are plotted in Figure $11.4.2$.

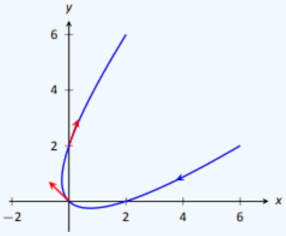


Figure $11.4.2$: Plotting unit tangent vectors in Example $11.4.2$.

Unit Normal Vector

Just as knowing the direction tangent to a path is important, knowing a direction orthogonal to a path is important. When dealing with real-valued functions, we defined the normal line at a point to the be the line through the point that was perpendicular to the tangent line at that point. We can do a similar thing with vector-valued functions. Given $\vec{\mathbf{r}}(t)$ in $\mathbb{R}^2$, we have 2 directions perpendicular to the tangent vector, as shown in Figure $11.4.3$. It is good to wonder "Is one of these two directions preferable over the other?"

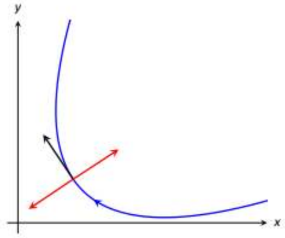


Figure $11.4.3$: Given a direction in the plane, there are always two directions orthogonal to it.

Given $\vec{\mathbf{r}}(t)$ in $\mathbb{R}^3$, there are infinite vectors orthogonal to the tangent vector at a given point. Again, we might wonder "Is one of these infinite choices preferable over the others? Is one of these the "right' choice?"

The answer in both $\mathbb{R}^2$ and $\mathbb{R}^3$ is "Yes, there is one vector that is not only preferable, it is the "right' one to choose." Recall Theorem 93, which states that if $\vec{\mathbf{r}}(t)$ has constant length, then $\vec{\mathbf{r}}(t)$ is orthogonal to $\vec{\mathbf{r}}\,'(t)$ for all t. We know $\vec{\mathbf{T}}(t)$, the unit tangent vector, has constant length. Therefore $\vec{\mathbf{T}}(t)$ is orthogonal to $\vec{\mathbf{T}}\,'(t)$.

We'll see that $\vec{\mathbf{T}}\,'(t)$ is more than just a convenient choice of vector that is orthogonal to $\vec{\mathbf{r}}\,'(t)$; rather, it is the "right" choice. Since all we care about is the direction, we define this newly found vector to be a unit vector.

Note: $\vec{\mathbf{T}}(t)$ is a unit vector, by definition. This *does not* imply that $\vec{\mathbf{T}}\,'(t)$ is also a unit vector.

Definition 11.4.2: Unit Normal Vector

Let $\vec{\mathbf{r}}(t)$ be a vector-valued function where the unit tangent vector, $\vec{\mathbf{T}}(t)$, is smooth on an open interval I. The **unit normal vector** $\vec{\mathbf{N}}(t)$ is

$$\vec{\mathbf{N}}(t)=\frac{1}{\|\vec{\mathbf{T}}\,'(t)\|}\vec{\mathbf{T}}\,'(t). \tag{11.4.4}$$

Example 11.4.3: Computing the unit normal vector

Let $\vec{\mathbf{r}}(t)=\langle 3\cos t, 3\sin t, 4t\rangle$ as in Example 11.4.1. Sketch both $\vec{\mathbf{T}}(\pi/2)$ and $\vec{\mathbf{N}}(\pi/2)$ with initial points at $\vec{\mathbf{r}}(\pi/2)$.

Solution

In Example 11.4.1, we found $\vec{\mathbf{T}}(t)=\langle(-3/5)\sin t,(3/5)\cos t,4/5\rangle$. Therefore

$$\vec{\mathbf{T}}\,'(t)=\langle-\frac{3}{5}\cos t,-\frac{3}{5}\sin t,0\rangle \quad \text{and} \quad \|\vec{\mathbf{T}}\,'(t)\|=\frac{3}{5}. \tag{11.4.5}$$

Thus

$$\vec{\mathbf{N}}(t)=\frac{\vec{\mathbf{T}}\,'(t)}{3/5}=\langle-\cos t,-\sin t,0\rangle. \tag{11.4.6}$$

We compute $\vec{\mathbf{T}}(\pi/2)=\langle-3/5,0,4/5\rangle$ and $\vec{\mathbf{N}}(\pi/2)=\langle0,-1,0\rangle$. These are sketched in Figure 11.4.4.

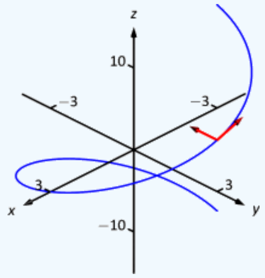


Figure 11.4.4: Plotting unit tangent and normal vectors in Example 11.4.3.

The previous example was once again "too nice." In general, the expression for $\vec{\mathbf{T}}(t)$ contains fractions of square-roots, hence the expression of $\vec{\mathbf{T}}\,'(t)$ is very messy. We demonstrate this in the next example.

Example 11.4.4: Computing the unit normal vector

Let $\vec{\mathbf{r}}(t)=\langle t^2-t, t^2+t\rangle$ as in Example 11.4.2. Find $\vec{\mathbf{N}}(t)$ and sketch $\vec{\mathbf{r}}(t)$ with the unit tangent and normal vectors at $t=-1,0$ and 1.

Solution

In Example 11.4.2, we found

$$\vec{\mathbf{T}}(t)=\langle\frac{2t-1}{\sqrt{8t^2+2}},\frac{2t+1}{\sqrt{8t^2+2}}\rangle. \tag{11.4.7}$$

Finding $\vec{\mathbf{T}}\,'(t)$ requires two applications of the Quotient Rule:

$$\begin{aligned}T'(t)&=\langle\frac{\sqrt{8t^2+2}(2)-(2t-1)\left(\frac{1}{2}(8t^2+2)^{-1/2}(16t)\right)}{8t^2+2},\\&\qquad\frac{\sqrt{8t^2+2}(2)-(2t+1)\left(\frac{1}{2}(8t^2+2)^{-1/2}(16t)\right)}{8t^2+2}\rangle\\&=\langle\frac{4(2t+1)}{(8t^2+2)^{3/2}},\frac{4(1-2t)}{(8t^2+2)^{3/2}}\rangle\end{aligned}$$

This is not a unit vector; to find $\vec{\mathbf{N}}(t)$, we need to divide $\vec{\mathbf{T}}\,'(t)$ by it's magnitude.

$$\begin{aligned}\|\vec{\mathbf{T}}\,'(t)\| &= \sqrt{\frac{16(2t+1)^2}{(8t^2+2)^3}+\frac{16(1-2t)^2}{(8t^2+2)^3}}\\ &= \sqrt{\frac{16(8t^2+2)}{(8t^2+2)^3}}\\ &= \frac{4}{8t^2+2}.\end{aligned}$$

Finally,

$$\begin{aligned}\vec{\mathbf{N}}(t) &= \frac{1}{4/(8t^2+2)}\langle\frac{4(2t+1)}{(8t^2+2)^{3/2}},\frac{4(1-2t)}{(8t^2+2)^{3/2}}\rangle\\ &= \langle\frac{2t+1}{\sqrt{8t^2+2}},-\frac{2t-1}{\sqrt{8t^2+2}}\rangle.\end{aligned}$$

Using this formula for $\vec{\mathbf{N}}(t)$, we compute the unit tangent and normal vectors for $t=-1,0$ and 1 and sketch them in Figure $11.4.5$.

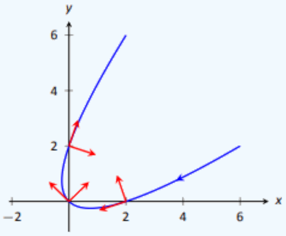


Figure $11.4.5$: Plotting unit tangent and normal vectors in Example $11.4.4$.

The final result for $\vec{\mathbf{N}}(t)$ in Example $11.4.4$ is suspiciously similar to $\vec{\mathbf{T}}(t)$. There is a clear reason for this. If $\vec{\mathbf{u}}=\langle u_1,u_2\rangle$ is a unit vector in $\mathbb{R}^2$, then the *only* unit vectors orthogonal to $\vec{\mathbf{u}}$ are $\langle -u_2,u_1\rangle$ and $\langle u_2,-u_1\rangle$. Given $\vec{\mathbf{T}}(t)$, we can quickly determine $\vec{\mathbf{N}}(t)$ if we know which term to multiply by (-1).

Consider again Figure 11.24, where we have plotted some unit tangent and normal vectors. Note how $\vec{\mathbf{N}}(t)$ always points "inside" the curve, or to the concave side of the curve. This is not a coincidence; this is true in general. Knowing the direction that $\vec{\mathbf{r}}(t)$ "turns" allows us to quickly find $\vec{\mathbf{N}}(t)$.

THEOREM 11.4.1: Unit Normal Vectors in $\mathbb{R}^2$

Let $\vec{\mathbf{r}}(t)$ be a vector-valued function in $\mathbb{R}^2$ where $\vec{\mathbf{T}}\,'(t)$ is smooth on an open interval I. Let t_0 be in I and $\vec{\mathbf{T}}(t_0)=\langle t_1,t_2\rangle$ Then $\vec{\mathbf{N}}(t_0)$ is either

$$\vec{\mathbf{N}}(t_0)=\langle -t_2,t_1\rangle \quad \text{or} \quad \vec{\mathbf{N}}(t_0)=\langle t_2,-t_1\rangle, \tag{11.4.8}$$

whichever is the vector that points to the concave side of the graph of $\vec{\mathbf{r}}$.

Application to Acceleration

Let $\vec{\mathbf{r}}(t)$ be a position function. It is a fact (stated later in Theorem $11.4.2$) that acceleration, \vecs a (t), lies in the plane defined by $\vec{\mathbf{T}}$ and $\vec{\mathbf{N}}$. That is, there are scalars a_{T} and a_{N} such that

$$\vec{\mathbf{a}}(t)=a_{\mathrm{T}}\vec{\mathbf{T}}(t)+a_{\mathrm{N}}\vec{\mathbf{N}}(t). \tag{11.4.9}$$

The scalar a_{T} measures "how much" acceleration is in the direction of travel, that is, it measures the component of acceleration that affects the speed. The scalar a_{N} measures "how much" acceleration is perpendicular to the direction of travel, that is, it measures the component of acceleration that affects the direction of travel.

We can find a_{T} using the orthogonal projection of $\vec{\mathbf{a}}(t)$ onto $\vec{\mathbf{T}}(t)$ (review Definition 59 in Section 10.3 if needed).

Recalling that since $\vec{\mathbf{T}}(t)$ is a unit vector, $\vec{\mathbf{T}}(t)\cdot\vec{\mathbf{T}}(t)=1$, so we have

$$\operatorname{proj}_{T(t)}\vec{\mathbf{a}}(t) = \frac{\vec{\mathbf{a}}(t)\cdot\vec{\mathbf{T}}(t)}{\vec{\mathbf{T}}(t)\cdot\vec{\mathbf{T}}(t)}\vec{\mathbf{T}}(t) = \underbrace{\left(\vec{\mathbf{a}}(t)\cdot\vec{\mathbf{T}}(t)\right)}_{a_{\mathrm{T}}}\vec{\mathbf{T}}(t). \tag{11.4.10}$$

Thus the amount of $\vec{\mathbf{a}}(t)$ in the direction of $\vec{\mathbf{T}}(t)$ is $a_{\mathrm{T}} = \vec{\mathbf{a}}(t)\cdot\vec{\mathbf{T}}(t)$. The same logic gives $a_{\mathrm{N}} = \vec{\mathbf{a}}(t)\cdot\vec{\mathbf{N}}(t)$.

While this is a fine way of computing a_{T}, there are simpler ways of finding a_{N} (as finding $\vec{\mathbf{N}}$ itself can be complicated). The following theorem gives alternate formulas for a_{T} and a_{N}.

Note: Keep in mind that both a_{T} and a_{N} are functions of t; that is, the scalar changes depending on t. It is convention to drop the "(t)" notation from $a_{\mathrm{T}}(t)$ and simply write a_{T}.

THEOREM 11.4.2: Acceleration in the Plane Defined by $\vec{\mathbf{T}}$ and $\vec{\mathbf{N}}$

Let $\vec{\mathbf{r}}(t)$ be a position function with acceleration $\vec{\mathbf{a}}(t)$ and unit tangent and normal vectors $\vec{\mathbf{T}}(t)$ and $\vec{\mathbf{N}}(t)$. Then $\vec{\mathbf{a}}(t)$ lies in the plane defined by $\vec{\mathbf{T}}(t)$ and $\vec{\mathbf{N}}(t)$; that is, there exists scalars a_{T} and a_{N} such that

$$\vec{\mathbf{a}}(t) = a_{\mathrm{T}}\vec{\mathbf{T}}(t) + a_{\mathrm{N}}\vec{\mathbf{N}}(t). \tag{11.4.11}$$

Moreover,

$$\begin{aligned} a_{\mathrm{T}} &= \vec{\mathbf{a}}(t)\cdot\vec{\mathbf{T}}(t) = \frac{d}{dt}\left(\|\vec{\mathbf{v}}(t)\|\right) \\ a_{\mathrm{N}} &= \vec{\mathbf{a}}(t)\cdot\vec{\mathbf{N}}(t) = \sqrt{\|\vec{\mathbf{a}}(t)\|^2 - a_{\mathrm{T}}^2} = \frac{\|\vec{\mathbf{a}}(t)\times\vec{\mathbf{v}}(t)\|}{\|\vec{\mathbf{v}}(t)\|} = \|\vec{\mathbf{v}}(t)\|\,\|\vec{\mathbf{T}}\,'(t)\| \end{aligned}$$

Note the second formula for a_{T}: $\frac{d}{dt}\left(\|\vec{\mathbf{v}}(t)\|\right)$. This measures the rate of change of speed, which again is the amount of acceleration in the direction of travel.

Example 11.4.5: Computing a_{T} and a_{N}

Let $\vec{\mathbf{r}}(t) = \langle 3\cos t, 3\sin t, 4t\rangle$ as in Examples 11.4.1 and 11.4.3. Find a_{T} and a_{N}.

Solution

The previous examples give $\vec{\mathbf{a}}(t) = \langle -3\cos t, -3\sin t, 0\rangle$ and

$$\vec{\mathbf{T}}(t) = \langle -\frac{3}{5}\sin t, \frac{3}{5}\cos t, \frac{4}{5}\rangle \quad \text{and} \quad \vec{\mathbf{N}}(t) = \langle -\cos t, -\sin t, 0\rangle. \tag{11.4.12}$$

We can find a_{T} and a_{N} directly with dot products:

$$\begin{aligned} a_{\mathrm{T}} &= \vec{\mathbf{a}}(t)\cdot\vec{\mathbf{T}}(t) = \frac{9}{5}\cos t\sin t - \frac{9}{5}\cos t\sin t + 0 = 0. \\ a_{\mathrm{N}} &= \vec{\mathbf{a}}(t)\cdot\vec{\mathbf{N}}(t) = 3\cos^2 t + 3\sin^2 t + 0 = 3. \end{aligned}$$

Thus $\vec{\mathbf{a}}(t) = 0\vec{\mathbf{T}}(t) + 3\vec{\mathbf{N}}(t) = 3\vec{\mathbf{N}}(t)$, which is clearly the case.

What is the practical interpretation of these numbers? $a_{\mathrm{T}} = 0$ means the object is moving at a constant speed, and hence all acceleration comes in the form of direction change.

Example 11.4.6: Computing a_{T} and a_{N}

Let $\vec{\mathbf{r}}(t) = \langle t^2 - t, t^2 + t\rangle$ as in Examples 11.4.2 and 11.4.4. Find a_{T} and a_{N}.

Solution

The previous examples give $\vec{\mathbf{a}}(t) = \langle 2, 2\rangle$ and

$$\vec{\mathbf{T}}(t) = \langle \frac{2t-1}{\sqrt{8t^2+2}}, \frac{2t+1}{\sqrt{8t^2+2}}\rangle$$

and

$$2t+1 \qquad 2t-1$$

$$\vec{\mathbf{N}}(t) = \langle \frac{}{\sqrt{8t^2+2}}, -\frac{}{\sqrt{8t^2+2}} \rangle.$$

While we can compute a_{N} using $\vec{\mathbf{N}}(t)$, we instead demonstrate using another formula from Theorem 11.4.2.

$$a_{\mathrm{T}} = \vec{\mathbf{a}}(t)\cdot\vec{\mathbf{T}}(t) = \frac{4t-2}{\sqrt{8t^2+2}} + \frac{4t+2}{\sqrt{8t^2+2}} = \frac{8t}{\sqrt{8t^2+2}}.$$

$$a_{\mathrm{N}} = \sqrt{\|\vec{\mathbf{a}}(t)\|^2 - a_{\mathrm{T}}^2} = \sqrt{8 - \left(\frac{8t}{\sqrt{8t^2+2}}\right)^2} = \frac{4}{\sqrt{8t^2+2}}$$

When $t = 2$, $a_{\mathrm{T}} = \dfrac{16}{\sqrt{34}} \approx 2.74$ and $a_{\mathrm{N}} = \dfrac{4}{\sqrt{34}} \approx 0.69$. We interpret this to mean that at $t = 2$, the particle is accelerating mostly by increasing speed, not by changing direction. As the path near $t = 2$ is relatively straight, this should make intuitive sense. Figure 11.4.6 gives a graph of the path for reference.

imageedit_18_5656399812.png

Figure 11.4.6: Graphing $\vec{\mathbf{r}}(t)$ in Example 11.4.6.

Contrast this with $t = 0$, where $a_{\mathrm{T}} = 0$ and $a_{\mathrm{N}} = 4/\sqrt{2} \approx 2.82$. Here the particle's speed is not changing and all acceleration is in the form of direction change.

Example 11.4.7: Analyzing projectile motion

A ball is thrown from a height of 240ft with an initial speed of 64ft/s and an angle of elevation of 30°. Find the position function $\vec{\mathbf{r}}(t)$ of the ball and analyze a_{T} and a_{N}.

Solution

Using Key Idea 53 of Section 11.3 we form the position function of the ball:

$$\vec{\mathbf{r}}(t) = \langle (64\cos 30^\circ)\,t, -16t^2 + (64\sin 30^\circ)\,t + 240\rangle, \tag{11.4.13}$$

which we plot in Figure 11.4.7.

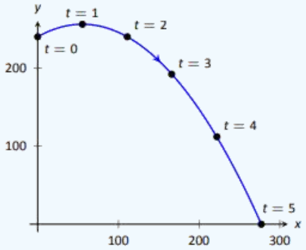


Figure 11.4.7: Plotting the position of a thrown ball, with 1s increments shown.

From this we find $\vec{\mathbf{v}}(t) = \langle 64\cos 30^\circ, -32t + 64\sin 30^\circ\rangle$ and $\vec{\mathbf{a}}(t) = \langle 0, -32\rangle$. Computing $\vec{\mathbf{T}}(t)$ is not difficult, and with some simplification we find

$$\vec{\mathbf{T}}(t) = \langle \frac{\sqrt{3}}{\sqrt{t^2-2t+4}}, \frac{1-t}{\sqrt{t^2-2t+4}} \rangle. \tag{11.4.14}$$

With $\vec{\mathbf{a}}(t)$ as simple as it is, finding a_{T} is also simple:

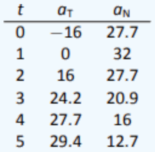
$$a_{\mathrm{T}} = \vec{\mathbf{a}}(t)\cdot\vec{\mathbf{T}}(t) = \frac{32t-32}{\sqrt{t^2-2t+4}}. \tag{11.4.15}$$

We choose to not find $\vec{\mathbf{N}}(t)$ and find a_{N} through the formula $a_{\mathrm{N}} = \sqrt{\|\vec{\mathbf{a}}(t)\|^2 - a_{\mathrm{T}}^2}$:

$$a_{\mathrm{N}} = \sqrt{32^2 - \left(\frac{32t-32}{\sqrt{t^2-2t+4}}\right)^2} = \frac{32\sqrt{3}}{\sqrt{t^2-2t+4}}. \tag{11.4.16}$$

Figure 11.4.8 gives a table of values of a_{T} and a_{N}. When $t = 0$, we see the ball's speed is decreasing; when $t = 1$ the speed of the ball is unchanged. This corresponds to the fact that at $t = 1$ the ball reaches its highest point.

After $t = 1$ we see that a_{N} is decreasing in value. This is because as the ball falls, it's path becomes straighter and most of the acceleration is in the form of speeding up the ball, and not in changing its direction.



t	a_T	a_N
0	−16	27.7
1	0	32
2	16	27.7
3	24.2	20.9
4	27.7	16
5	29.4	12.7

Figure 11.4.8: A table of values of a_T and a_N in Example 11.4.7.

Our understanding of the unit tangent and normal vectors is aiding our understanding of motion. The work in Example 11.4.7 gave quantitative analysis of what we intuitively knew.

The next section provides two more important steps towards this analysis. We currently describe position only in terms of time. In everyday life, though, we often describe position in terms of distance ("The gas station is about 2 miles ahead, on the left."). The *arc length parameter* allows us to reference position in terms of distance traveled.

We also intuitively know that some paths are straighter than others - and some are "curvier" than others, but we lack a measurement of "curviness." The arc length parameter provides a way for us to compute *curvature*, a quantitative measurement of how curvy a curve is.

This page titled 11.4: Unit Tangent and Normal Vectors is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

11.5: The Arc Length Parameter and Curvature

In normal conversation we describe position in terms of both *time* and *distance*. For instance, imagine driving to visit a friend. If she calls and asks where you are, you might answer "I am 20 minutes from your house," or you might say "I am 10 miles from your house." Both answers provide your friend with a general idea of where you are.

Currently, our vector-valued functions have defined points with a parameter t, which we often take to represent time. Consider Figure $11.5.1a$, where $\vec{\mathbf{r}}(t) = \langle t^2 - t, t^2 + t \rangle$ is graphed and the points corresponding to $t = 0,\ 1$ and 2 are shown. Note how the arc length between $t = 0$ and $t = 1$ is smaller than the arc length between $t = 1$ and $t = 2$; if the parameter t is time and $\vec{\mathbf{r}}$ is position, we can say that the particle traveled faster on $[1, 2]$ than on $[0, 1]$.

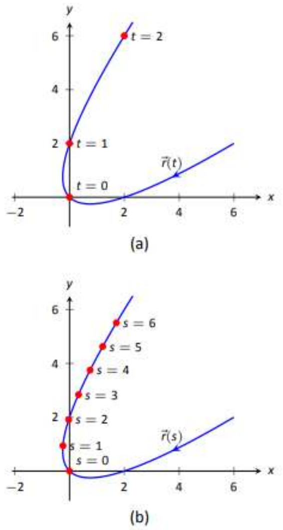


Figure $11.5.1$: Introducing the arc length parameter.

Now consider Figure $11.5.1b$, where the same graph is parametrized by a different variable s. Points corresponding to $s = 0$ through $s = 6$ are plotted. The arc length of the graph between each adjacent pair of points is 1. We can view this parameter s as distance; that is, the arc length of the graph from $s = 0$ to $s = 3$ is 3, the arc length from $s = 2$ to $s = 6$ is 4, etc. If one wants to find the point 2.5 units from an initial location (i.e., $s = 0$), one would compute $\vec{\mathbf{r}}(2.5)$. This parameter s is very useful, and is called the **arc length parameter**.

How do we find the arc length parameter?

Start with any parametrization of $\vec{\mathbf{r}}$. We can compute the arc length of the graph of $\vec{\mathbf{r}}$ on the interval $[0, t]$ with

$$\text{arc length } = \int_0^t \|\vec{\mathbf{r}}\,'(u)\| du. \tag{11.5.1}$$

We can turn this into a function: as t varies, we find the arc length s from 0 to t. This function is

$$s(t) = \int_0^t \|\vec{\mathbf{r}}\,'(u)\| du. \tag{11.5.2}$$

This establishes a relationship between s and t. Knowing this relationship explicitly, we can rewrite $\vec{\mathbf{r}}(t)$ as a function of s: $\vec{\mathbf{r}}(s)$. We demonstrate this in an example.

Example $11.5.1$: Finding the arc length parameter

Let $\vec{\mathbf{r}}(t) = \langle 3t - 1, 4t + 2 \rangle$. Parametrize $\vec{\mathbf{r}}$ with the arc length parameter s.}

Solution

Using Equation $11.5.2$, we write

$$s(t) = \int_0^t \|\vec{\mathbf{r}}\,'(u)\| du.$$

We can integrate this, explicitly finding a relationship between s and t:

$$\begin{aligned} s(t) &= \int_0^t \|\vec{\mathbf{r}}\,'(u)\| du \\ &= \int_0^t \sqrt{3^2+4^2} du \\ &= \int_0^t 5du \\ &= 5t. \end{aligned}$$

Since $s=5t$, we can write $t=s/5$ and replace t in $\vec{\mathbf{r}}(t)$ with $s/5$:

$$\vec{\mathbf{r}}(s) = \langle 3(s/5)-1, 4(s/5)+2\rangle = \langle \frac{3}{5}s-1, \frac{4}{5}s+2\rangle.$$

Clearly, as shown in Figure $11.5.2$, the graph of $\vec{\mathbf{r}}$ is a line, where $t=0$ corresponds to the point $(-1,2)$. What point on the line is 2 units away from this initial point? We find it with $s(2)=\langle 1/5, 18/5\rangle$.

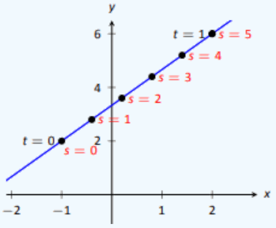


Figure $11.5.2$: Graphing $vecr$ in Example $11.5.1$.

Is the point $(1/5, 18/5)$ really 2 units away from $(-1,2)$? We use the Distance Formula to check:

$$\begin{aligned} d &= \sqrt{\left(\frac{1}{5}-(-1)\right)^2+\left(\frac{18}{5}-2\right)^2} \\ &= \sqrt{\frac{36}{25}+\frac{64}{25}} \\ &= \sqrt{4}=2. \end{aligned}$$

Yes, $s(2)$ is indeed 2 units away, in the direction of travel, from the initial point.

Things worked out very nicely in Example $11.5.1$; we were able to establish directly that $s=5t$. Usually, the arc length parameter is much more difficult to describe in terms of t, a result of integrating a square-root. There are a number of things that we can learn about the arc length parameter from Equation 11.5.2, though, that are incredibly useful.

First, take the derivative of s with respect to t. The Fundamental Theorem of Calculus (see Theorem 39) states that

$$\frac{ds}{dt} = s\,'(t) = \|\vec{\mathbf{r}}\,'(t)\|. \tag{11.5.3}$$

Letting t represent time and $\vec{\mathbf{r}}(t)$ represent position, we see that the rate of change of s with respect to t is speed; that is, the rate of change of "distance traveled" is speed, which should match our intuition.

The Chain Rule states that

$$\begin{aligned} \frac{d\vec{\mathbf{r}}}{dt} &= \frac{d\vec{\mathbf{r}}}{ds}\cdot\frac{ds}{dt} \\ \vec{\mathbf{r}}\,'(t) &= \vec{\mathbf{r}}\,'(s)\cdot\|\vec{\mathbf{r}}\,'(t)\|. \end{aligned}$$

Solving for $\vec{\mathbf{r}}\,'(s)$, we have

$$\vec{\mathbf{r}}\,'(s) = \frac{\vec{\mathbf{r}}\,'(t)}{\|\vec{\mathbf{r}}\,'(t)\|} = \vec{\mathbf{T}}(t), \tag{11.5.4}$$

where $\vec{\mathbf{T}}(t)$ is the unit tangent vector. Equation 11.5.4 is often misinterpreted, as one is tempted to think it states $\vec{\mathbf{r}}\,'(t) = \vec{\mathbf{T}}(t)$, but there is a big difference between $\vec{\mathbf{r}}\,'(s)$ and $\vec{\mathbf{r}}\,'(t)$. The key to take from it is that $\vec{\mathbf{r}}\,'(s)$ is a unit vector. In fact, the following theorem states that this characterizes the arc length parameter.

theorem 99: Arc Length Parameter

Let $\vec{\mathbf{r}}(s)$ be a vector-valued function. The parameter s is the arc length parameter if, and only if, $\|\vec{\mathbf{r}}\,'(s)\| = 1$.

Curvature

Consider points A and B on the curve graphed in Figure $11.5.3a$. One can readily argue that the curve curves more sharply at A than at B. It is useful to use a number to describe how sharply the curve bends; that number is the **curvature** of the curve.

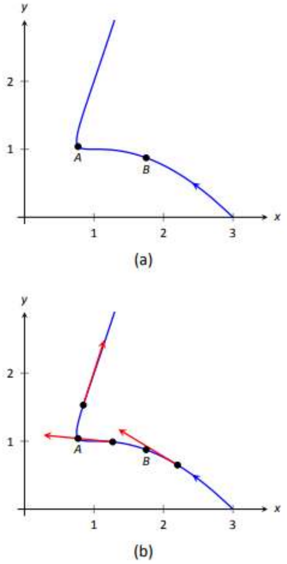


Figure $11.5.3$: Establishing the concept of curvature.

We derive this number in the following way. Consider Figure $11.5.3b$, where unit tangent vectors are graphed around points A and B. Notice how the direction of the unit tangent vector changes quite a bit near A, whereas it does not change as much around B. This leads to an important concept: measuring the rate of change of the unit tangent vector with respect to arc length gives us a measurement of curvature.

Definition 11.5.1: Curvature

Let $\vec{\mathbf{r}}(s)$ be a vector-valued function where s is the arc length parameter. The curvature κ of the graph of $\vec{\mathbf{r}}(s)$ is

$$\kappa = \left\|\frac{d\vec{\mathbf{T}}}{ds}\right\| = \|\vec{\mathbf{T}}\,'(s)\|. \tag{11.5.5}$$

If $\vec{\mathbf{r}}(s)$ is parametrized by the arc length parameter, then

$$\vec{\mathbf{T}}(s) = \frac{\vec{\mathbf{r}}\,'(s)}{\|\vec{\mathbf{r}}\,'(s)\|} \quad \text{and} \quad \vec{\mathbf{N}}(s) = \frac{\vec{\mathbf{T}}\,'(s)}{\|\vec{\mathbf{T}}\,'(s)\|}. \tag{11.5.6}$$

Having defined $\|\vec{\mathbf{T}}\,'(s)\| = \kappa$, we can rewrite the second equation as

$$\vec{\mathbf{T}}\,'(s) = \kappa\vec{\mathbf{N}}(s). \tag{11.5.7}$$

We already knew that $\vec{\mathbf{T}}\,'(s)$ is in the same direction as $\vec{\mathbf{N}}(s)$; that is, we can think of $\vec{\mathbf{T}}(s)$ as being "pulled" in the direction of $\vec{\mathbf{N}}(s)$. How "hard" is it being pulled? By a factor of κ. When the curvature is large, $\vec{\mathbf{T}}(s)$ is being "pulled hard" and the direction of $\vec{\mathbf{T}}(s)$ changes rapidly. When κ is small, $T(s)$ is not being pulled hard and hence its direction is not changing rapidly.

We use Definition $11.5.1$ to find the curvature of the line in Example $11.5.2$.

Example 11.5.2: Finding the curvature of a line

Use Definition 11.5.1 to find the curvature of $\vec{\mathbf{r}}(t) = \langle 3t-1, 4t+2\rangle$.

Solution

In Example ???, we found that the arc length parameter was defined by $s = 5t$, so $\vec{\mathbf{r}}(s) = \langle 3t/5-1, 4t/5+2\rangle$ parametrized $\vec{\mathbf{r}}$ with the arc length parameter. To find κ, we need to find $\vec{\mathbf{T}}\,'(s)$.

$$\begin{aligned}\vec{\mathbf{T}}(s) &= \vec{\mathbf{r}}\,'(s) \quad \text{(recall this is a unit vector)} \\ &= \langle 3/5, 4/5\rangle.\end{aligned}$$

Therefore

$$\vec{\mathbf{T}}\,'(s) = \langle 0, 0\rangle$$

and

$$\kappa = \|\vec{\mathbf{T}}\,'(s)\| = 0.$$

It probably comes as no surprise that the curvature of a line is 0. (How "curvy" is a line? It is not curvy at all.)

While the definition of curvature is a beautiful mathematical concept, it is nearly impossible to use most of the time; writing $\vec{\mathbf{r}}$ in terms of the arc length parameter is generally very hard. Fortunately, there are other methods of calculating this value that are much easier. There is a tradeoff: the definition is "easy" to understand though hard to compute, whereas these other formulas are easy to compute though it may be hard to understand why they work.

theorem 100: Formulas for Curvature

Let C be a smooth curve on an open interval I in the plane or in space.

1. If C is defined by $y = f(x)$, then

$$\kappa = \frac{|f''(x)|}{\left(1+\left(f'(x)\right)^2\right)^{3/2}}. \tag{11.5.8}$$

2. If C is defined as a vector-valued function in the plane, $\vec{\mathbf{r}}(t) = \langle x(t), y(t)\rangle$, then

$$\kappa = \frac{|x'y''-x''y'|}{\left((x')^2+(y')^2\right)^{3/2}}. \tag{11.5.9}$$

3. If C is defined in space by a vector-valued function $\vec{\mathbf{r}}(t)$, then

$$\kappa = \frac{\|\vec{\mathbf{T}}\,'(t)\|}{\|\vec{\mathbf{r}}\,'(t)\|} = \frac{\|\vec{\mathbf{r}}\,'(t)\times\vec{\mathbf{r}}\,''(t)\|}{\|\vec{\mathbf{r}}\,'(t)\|^3} = \frac{\vec{\mathbf{a}}(t)\cdot\vec{\mathbf{N}}(t)}{\|\vec{\mathbf{v}}(t)\|^2}. \tag{11.5.10}$$

We practice using these formulas.

Example 11.5.3: Finding the curvature of a circle

Find the curvature of a circle with radius r, defined by $\vec{\mathbf{c}}(t) = \langle r\cos t, r\sin t\rangle$.

Solution

Before we start, we should expect the curvature of a circle to be constant, and not dependent on t. (Why?)

We compute κ using the second part of Theorem 100.

$$\begin{aligned}\kappa &= \frac{|(-r\sin t)(-r\sin t)-(-r\cos t)(r\cos t)|}{\left((-r\sin t)^2+(r\cos t)^2\right)^{3/2}}\\ &= \frac{r^2(\sin^2 t+\cos^2 t)}{\left(r^2(\sin^2 t+\cos^2 t)\right)^{3/2}}\\ &= \frac{r^2}{r^3}=\frac{1}{r}.\end{aligned}$$

We have found that a circle with radius r has curvature $\kappa = 1/r$.

Example $11.5.3$ gives a great result. Before this example, if we were told "The curve has a curvature of 5 at point A," we would have no idea what this really meant. Is 5 "big" - does is correspond to a really sharp turn, or a not-so-sharp turn? Now we can think of 5 in terms of a circle with radius 1/5. Knowing the units (inches vs. miles, for instance) allows us to determine how sharply the curve is curving.

Let a point P on a smooth curve C be given, and let κ be the curvature of the curve at P. A circle that:

- passes through P,
- lies on the concave side of C,
- has a common tangent line as C at P and
- has radius $r = 1/\kappa$ (hence has curvature κ)

is the **osculating circle**, or **circle of curvature**, to C at P, and r is the **radius of curvature**. Figure $11.5.4$ shows the graph of the curve seen earlier in Figure 11.30 and its osculating circles at A and B. A sharp turn corresponds to a circle with a small radius; a gradual turn corresponds to a circle with a large radius. Being able to think of curvature in terms of the radius of a circle is very useful.

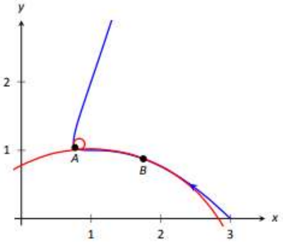


Figure 11.30

Example $11.5.4$: Finding curvature

Find the curvature of the parabola defined by $y = x^2$ at the vertex and at $x = 1$.

Solution

We use the first formula found in Theorem 100.

$$\begin{aligned}\kappa(x) &= \frac{|2|}{\left(1+(2x)^2\right)^{3/2}}\\ &= \frac{2}{\left(1+4x^2\right)^{3/2}}.\end{aligned}$$

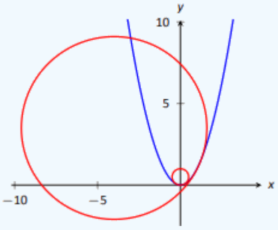


Figure $11.5.5$: Examining the curvature of $y = x^2$.

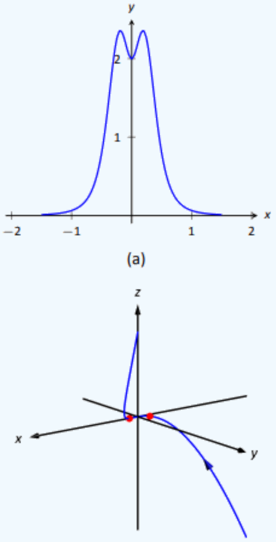
At the vertex ($x=0$), the curvature is $\kappa=2$. At $x=1$, the curvature is $\kappa=2/(5)^{3/2}\approx 0.179$. So at $x=0$, the curvature of $y=x^2$ is that of a circle of radius $1/2$; at $x=1$, the curvature is that of a circle with radius $\approx 1/0.179\approx 5.59$. This is illustrated in Figure $11.5.5$. At $x=3$, the curvature is 0.009; the graph is nearly straight as the curvature is very close to 0.

Example $11.5.5$: Finding curvature

Find where the curvature of $\vec{\textbf{r}}(t)=\langle t,t^2,2t^3\rangle$ is maximized.

Solution
We use the third formula in Theorem 100 as $\vec{\textbf{r}}(t)$ is defined in space. We leave it to the reader to verify that

$$\begin{aligned} \vec{\textbf{r}}\,'(t)=\langle 1,2t,6t^2\rangle, \quad \vec{\textbf{r}}\,''(t)=\langle 0,2,12t\rangle, \\ \text{and} \quad \vec{\textbf{r}}\,'(t)\times\vec{\textbf{r}}\,''(t)=\langle 12t^2,-12t,2\rangle. \end{aligned} \tag{11.5.11}$$



(a)

(b)

Figure $11.5.6$: Understanding the curvature of a curve in space.

Thus

$$\begin{aligned} \kappa(t) &= \frac{\|\vec{\textbf{r}}\,'(t)\times\vec{\textbf{r}}\,''(t)\|}{\|\vec{\textbf{r}}\,'(t)\|^3} \\ &= \frac{\|\langle 12t^2,-12t,2\rangle\|}{\|\langle 1,2t,6t^2\rangle\|^3} \\ &= \frac{\sqrt{144t^4+144t^2+4}}{\left(\sqrt{1+4t^2+36t^4}\right)^3} \end{aligned}$$

While this is not a particularly "nice" formula, it does explicitly tell us what the curvature is at a given t value. To maximize $\kappa(t)$, we should solve $\kappa'(t)=0$ for t. This is doable, but *very* time consuming. Instead, consider the graph of $\kappa(t)$ as given in Figure $11.5.6a$. We see that κ is maximized at two t values; using a numerical solver, we find these values are $t\approx\pm 0.189$. In Figure $11.5.1b$ we graph $\vec{\textbf{r}}(t)$ and indicate the points where curvature is maximized.

Curvature and Motion

Let $\vec{\textbf{r}}(t)$ be a position function of an object, with velocity $\vec{\textbf{v}}(t)=\vec{\textbf{r}}\,'(t)$ and acceleration $\vec{\textbf{a}}(t)=\vec{\textbf{r}}\,''(t)$. In Section 11.4 we established that acceleration is in the plane formed by $\vec{\textbf{T}}(t)$ and $\vec{\textbf{N}}(t)$, and that we can find scalars a_{T} and a_{N} such that

$$\vec{\textbf{a}}(t)=a_{\text{T}}\vec{\textbf{T}}(t)+a_{\text{N}}\vec{\textbf{N}}(t). \tag{11.5.12}$$

Theorem 98 gives formulas for a_{T} and a_{N}:

$$a_{\text{T}}=\frac{d}{dt}\left(\|\vec{\textbf{v}}(t)\|\right) \quad \text{and} \quad a_{\text{N}}=\frac{\|\vec{\textbf{v}}(t)\times\vec{\textbf{a}}(t)\|}{\|\vec{\textbf{v}}(t)\|}. \tag{11.5.13}$$

We understood that the amount of acceleration in the direction of $\vec{\mathbf{T}}$ relates only to how the speed of the object is changing, and that the amount of acceleration in the direction of $\vec{\mathbf{N}}$ relates to how the direction of travel of the object is changing. (That is, if the object travels at constant speed, $a_{\mathrm{T}}=0$; if the object travels in a constant direction, $a_{\mathrm{N}}=0$.)

In Equation 11.5.3 at the beginning of this section, we found $s'(t)=\|\vec{\mathbf{v}}(t)\|$. We can combine this fact with the above formula for a_{T} to write

$$a_{\mathrm{T}}=\frac{d}{dt}\Big(\|\vec{\mathbf{v}}(t)\|\Big)=\frac{d}{dt}\Big(s'(t)\Big)=s''(t). \tag{11.5.14}$$

Since $s'(t)$ is speed, $s''(t)$ is the rate at which speed is changing with respect to time. We see once more that the component of acceleration in the direction of travel relates only to speed, not to a change in direction.

Now compare the formula for a_{N} above to the formula for curvature in Theorem 100:

$$a_{\mathrm{N}}=\frac{\|\vec{\mathbf{v}}(t)\times\vec{\mathbf{a}}(t)\|}{\|\vec{\mathbf{v}}(t)\|} \quad \text{and} \quad \kappa=\frac{\|\vec{\mathbf{r}}'(t)\times\vec{\mathbf{r}}''(t)\|}{\|\vec{\mathbf{r}}'(t)\|^3}=\frac{\|\vec{\mathbf{v}}(t)\times\vec{\mathbf{a}}(t)\|}{\|\vec{\mathbf{v}}(t)\|^3}. \tag{11.5.15}$$

Thus

$$\begin{aligned} a_{\mathrm{N}} &= \kappa\|\vec{\mathbf{v}}(t)\|^2 \\ &= \kappa\Big(s'(t)\Big)^2 \end{aligned} \tag{11.5.16}$$

This last equation shows that the component of acceleration that changes the object's direction is dependent on two things: the curvature of the path and the speed of the object.

Imagine driving a car in a clockwise circle. You will naturally feel a force pushing you towards the door (more accurately, the door is pushing you as the car is turning and you want to travel in a straight line). If you keep the radius of the circle constant but speed up (i.e., increasing $s'(t)$), the door pushes harder against you (a_{N} has increased). If you keep your speed constant but tighten the turn (i.e., increase κ), once again the door will push harder against you.

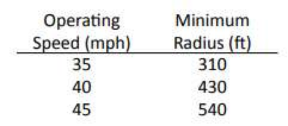
Putting our new formulas for a_{T} and a_{N} together, we have

$$\vec{\mathbf{a}}(t)=s''(t)\vec{\mathbf{T}}(t)+\kappa\|\vec{\mathbf{v}}(t)\|^2\vec{\mathbf{N}}(t). \tag{11.5.17}$$

This is not a particularly practical way of finding a_{T} and a_{N}, but it reveals some great concepts about how acceleration interacts with speed and the shape of a curve.

Example 11.5.6: Curvature and road design

The minimum radius of the curve in a highway cloverleaf is determined by the operating speed, as given in the table in Figure 11.5.7. For each curve and speed, compute a_{N}.



Operating Speed (mph)	Minimum Radius (ft)
35	310
40	430
45	540

Figure 11.5.7: Operating speed and minimum radius in highway cloverleaf design.

Solution

Using Equation 11.5.16, we can compute the acceleration normal to the curve in each case. We start by converting each speed from "miles per hour" to "feet per second" by multiplying by $5280/3600$.

$$\begin{aligned} 35\text{mph},\ 310\text{ft} \ &\Rightarrow 51.33\text{ft/s}, \quad \kappa=1/310 \\ a_{\mathrm{N}} &= \kappa\ \|\vec{\mathbf{v}}(t)\|^2 \\ &= \frac{1}{310}\Big(51.33\Big)^2 \\ &= 8.50\text{ft/s}^2. \end{aligned}$$

$$\text{40mph, 430ft} \ \Rightarrow 58.67\text{ft/s}, \quad \kappa = 1/430$$

$$\begin{aligned} a_{\mathrm{N}} &= \frac{1}{430}(58.67)^2 \\ &= 8.00\text{ft/s}^2. \end{aligned}$$

$$\text{45mph,540ft} \ \Rightarrow 66\text{ft/s}, \quad \kappa = 1/540$$

$$\begin{aligned} a_{\mathrm{N}} &= \frac{1}{540}(66)^2 \\ &= 8.07\text{ft/s}^2. \end{aligned}$$

Note that each acceleration is similar; this is by design. Considering the classic "Force $=$ mass $\times$ acceleration" formula, this acceleration must be kept small in order for the tires of a vehicle to keep a "grip" on the road. If one travels on a turn of radius 310ft at a rate of 50mph, the acceleration is double, at 17.35ft/s^2. If the acceleration is too high, the frictional force created by the tires may not be enough to keep the car from sliding. Civil engineers routinely compute a "safe" design speed, then subtract 5-10mph to create the posted speed limit for additional safety.

We end this chapter with a reflection on what we've covered. We started with vector-valued functions, which may have seemed at the time to be just another way of writing parametric equations. However, we have seen that the vector perspective has given us great insight into the behavior of functions and the study of motion. Vector-valued position functions convey displacement, distance traveled, speed, velocity, acceleration and curvature information, each of which has great importance in science and engineering.

This page titled 11.5: The Arc Length Parameter and Curvature is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

11.E: Applications of Vector Valued Functions (Exercises)

11.1: Vector-Valued Functions

Terms and Concepts

1. Vector-valued functions are closely related to _______ ________ of graphs.

2. When sketching vector-valued functions, technically one isn't graphing points, but rather _______.

3. It can be useful to think of _______ as a vector that points from a starting position to an ending position.

Problems

In Exercises 4-11, sketch the vector-valued function on the given interval.

4. $\vec{r}(t)=\langle t^2,t^2-1\rangle,\text{ for }-2\leq t\leq 2.$

5. $\vec{r}(t)=\langle t^2,t^3\rangle,\text{ for }-2\leq t\leq 2.$

6. $\vec{r}(t)=\langle 1/t,t/t^2\rangle,\text{ for }-2\leq t\leq 2.$

7. $\vec{r}(t)=\langle \frac{1}{10}t^2,\sin t\rangle,\text{ for }-2\pi\leq t\leq 2\pi.$

8. $\vec{r}(t)=\langle \frac{1}{10}t^2,\sin t\rangle,\text{ for }-2\pi\leq t\leq 2\pi.$

9. $\vec{r}(t)=\langle 3\sin(\pi t),2\cos(\pi t)\rangle,\text{ on }[0,2].$

10. $\vec{r}(t)=\langle 3\cos t,2\sin(2t)\rangle,\text{ on }[0,2\pi].$

11. $\vec{r}(t)=\langle 2\sec t,\tan t\rangle,\text{ on }[-\pi,\pi].$

In Exercises 12-15, sketch the vector-valued function on the given interval in $\mathbb{R}^3$. Technology may be useful in creating the sketch.

12. $\vec{r}(t)=\langle 2\cos t,t,2\sin t\rangle,\text{ on }[0,2\pi].$

13. $\vec{r}(t)=\langle 3\cos t,\sin t,t/\pi\rangle,\text{ on }[0,2\pi].$

14. $\vec{r}(t)=\langle \cos t,\sin t,\sin t\rangle,\text{ on }[0,2\pi].$

15. $\vec{r}(t)=\langle \cos t,\sin t,\sin(2t)\rangle,\text{ on }[0,2\pi].$

In Exercises 16-19, find $\|\vec{r}(t)\|$.

16. $\vec{r}(t)=\langle t,t^2\rangle.$

17. $\vec{r}(t)=\langle 5\cos t,3\sin t\rangle.$

18. $\vec{r}(t)=\langle 2\cos t,2\sin t,t\rangle.$

19. $\vec{r}(t)=\langle \cos t,t,t^2\rangle.$

In Exercises 20-27, create a vector-valued function whose graph matches the given description.

20. A circle of radius 2, centered $(1,2)$, traced counter-clockwise once on $[0,2\pi]$.

21. A circle of radius 3, centered $(5,5)$, traced clockwise once on $[0,2\pi]$.

22. An ellipse, centered at $(0,0)$ with vertical major axis length 10 and minor axis of length 3, traced once counter-clockwise on $[0,2\pi]$.

23. An ellipse, centered at $(3,-2)$ with horizontal major axis length 6 and minor axis of length 4, traced once clockwise on $[0,2\pi]$.

24. A line through $(2,3)$ with a slope of 5.

25. A line through $(1,5)$ with a slope of -1/2.

26. A vertically oriented helix with radius of 2 that starts at $(2,0,0)$ and ends at $(2,0,4\pi)$ after 1 revolution on $[0,2\pi]$.

27. A vertically oriented helix with radius of 3 that starts at $(3,0,0)$ and ends at $(3,0,3)$ after 2 revolutions on $[0,1]$.

In Exercises 28-31, find the average rate of change of $\vec{r}(t)$ on the given interval.

28. $\vec{r}(t)=\langle t,t^2\rangle$ on $[-2,2]$.

29. $\vec{r}(t)=\langle t,t+\sin t\rangle$ on $[0,2\pi]$.

30. $\vec{r}(t)=\langle 3\cos t,2\sin t,t\rangle$ on $[0,2\pi]$.

31. $\vec{r}(t)=\langle t,t^2,t^3\rangle$ on $[-1,3]$.

11.2: Calculus and Vector-Valued Functions

Terms and Concepts

1. Limits, derivatives and integrals of vector-valued functions are all evaluated ________ -wise.

2. The definite integral of a rate of change function gives ________.

3. Why is it generally not useful to graph both $\vec{r}(t)$ and $\vec{r}\,'(t)$ on the same axes?

Problems

In Exercises 4-7, evaluate the given limit.

4. $\lim\limits_{t\to 5}\langle 2t+1,3t^2-1,\sin t\rangle$

5. $\lim\limits_{t\to 3}\langle e^t,\frac{t^2-9}{t+3}\rangle$

6. $\lim\limits_{t\to 0}\langle \frac{t}{\sin t},(1+t)^{\frac{1}{t}}\rangle$

7. $\lim\limits_{h\to 0}\frac{\vec{r}(t+h)-\vec{r}(t)}{h}$, where $\vec{r}(t)=\langle t^2,t,1\rangle$

In Exercises 8-9, identify the interval(s) on which $\vec{r}(t)$ is continuous.

8. $\vec{r}(t)=\langle t^2,1/t\rangle$

9. $\vec{r}(t)=\langle \cos t,e^t,\ln t\rangle$

In Exercises 10-14, find the derivative of the given function.

10. $\vec{r}(t)=\langle \cos t,e^t,\ln t\rangle$

11. $\vec{r}(t)=\left\langle \frac{1}{t},\frac{2t-1}{3t+1},\tan t\right\rangle$

12. $\vec{r}(t)=(t^2)\langle \sin t,2t+5\rangle$

13. $\vec{r}(t)=\langle t^2+1,t-1\rangle\cdot\langle \sin t,2t+5\rangle$

14. $\vec{r}(t)=\langle t^2+1,t-1,1\rangle\times\langle \sin t,2t+5,1\rangle$

In Exercises 15-18, find $\vec{r}\,'(t)$. Sketch $\vec{r}(t)$ and $\vec{r}\,'(1)$, with the initial point of $\vec{r}\,'(1)$ at $\vec{r}(1)$.

15. $\vec{r}(t)=\langle t^2+t,t^2-t\rangle$

16. $\vec{r}(t)=\langle t^2-2t+2,t^3-3t^2+2t\rangle$

17. $\vec{r}(t)=\langle t^2+1,t^3-t\rangle$

18. $\vec{r}(t)=\langle t^2-4t+5,t^3-6t^2+11t-6\rangle$

In Exercises 19-22, give the equation of the line tangent to the graph of $\vec{r}(t)$ at the given *t* value.

19. $\vec{r}(t)=\langle t^2+t,t^2-t\rangle$ at $t=1$.

20. $\vec{r}(t)=\langle 3\cos t,\sin t\rangle$ at $t=\pi/4$.

21. $\vec{r}(t)=\langle 3\cos t,3\sin t,t\rangle$ at $t=\pi$.

22. $\vec{r}(t)=\langle e^t,\tan t,t\rangle$ at $t=0$.

In Exercises 23-26, find the value(s) of *t* for which $\vec{r}(t)$ is not smooth.

23. $\vec{r}(t)=\langle \cos t,\sin t-t\rangle$

24. $\vec{r}(t)=\langle t^2-2t+1,t^3+t^2-5t+3\rangle$

25. $\vec{r}(t)=\langle\cos t-\sin t-\cos t,\cos(4t)\rangle$

26. $\vec{r}(t)=\langle t^3-3t+2,-\cos(\pi t),\sin^2(\pi t)\rangle$

Exercises 27-29 ask you to verify parts of Theorem 92. In each let $f(t)=t^3,\vec{r}(t)=\langle t^2,t-1,1\rangle$ **and** $\vec{s}(t)=\langle\sin t,e^t,t\rangle$**. Compute the various derivatives as indicated.**

27. Simplify $f(t)\vec{r}(t)$, then find its derivative; show this is the same as $f'(t)\vec{r}'(t)+f(t)\vec{r}'(t)$.

28. Simplify $\vec{r}(t)\cdot\vec{s}(t)$, then find its derivative; show this is the same as $\vec{r}'(t)\cdot\vec{s}'(t)+\vec{r}(t)\cdot\vec{s}'(t)$.

29. Simplify $\vec{r}(t)\times\vec{s}(t)$, then find its derivative; show this is the same as $\vec{r}'(t)\times\vec{s}'(t)+\vec{r}(t)\times\vec{s}'(t)$.

In Exercises 30-33, evaluate the given definite or indefinite integral.

30. $\int\langle t^3,\cos t,te^t\rangle\,dt$

31. $\int\left\langle\frac{1}{1+t^2},\sec^2 t\right\rangle\,dt$

32. $\int_0^\pi\langle-\sin t,\cos t\rangle\,dt$

33. $\int_{-2}^2\langle 2t+1,2t-1\rangle\,dt$

In Exercises 34-37, solve the given initial value problems.

34. Find $\vec{r}(t)$, given that $\vec{r}'(t)=\langle t,\sin t\rangle$ and $\vec{r}(0)=\langle 2,2\rangle$.

35. Find $\vec{r}(t)$, given that $\vec{r}'(t)=\langle 1,(t+1),\tan t\rangle$ and $\vec{r}(0)=\langle 1,2\rangle$.

36. Find $\vec{r}(t)$, given that $\vec{r}''(t)=\langle t^2,t,1\rangle,\vec{r}'(0)=\langle 1,2,3\rangle\text{ and }\vec{r}(0)=\langle 4,5,6\rangle$.

37. Find $\vec{r}(t)$, given that $\vec{r}''(t)=\langle\cos t,\sin t,e^t\rangle,\vec{r}'(0)=\langle 0,0,0\rangle\text{ and }\vec{r}(0)=\langle 0,0,0\rangle$.

In Exercises 38-41, find the arc length of $\vec{r}(t)$ on the indicated interval.

38. $\vec{r}(t)=\langle 2\cos t,2\sin t,3t\rangle\text{ on }[0,2\pi]$

39. $\vec{r}(t)=\langle 5\cos t,3\sin t,4\sin t\rangle\text{ on }[0,2\pi]$.

40. $\vec{r}(t)=\langle t^3,t^2,t^3\rangle\text{ on }[0,1]$.

41. $\vec{r}(t)=\langle e^{-t}\cos t,e^{-t}\sin t\rangle\text{ on }[0,1]$.

42. Prove Theorem 93; that is, show if $\vec{r}(t)$ has constant length and is differentiable, then $\vec{r}(t)\cdot\vec{r}'(t)=0$. (Hint: use the Product Rule to compute $\frac{d}{dt}(\vec{r}(t)\cdot\vec{r}(t))$.)

11.3: The Calculus of Motion

Terms and Concepts

1. How is velocity different from speed?

2. What is the difference between displacement and distance traveled?

3. What is the difference between average velocity and average speed?

4. Distance traveled is the same as ______ _______, just viewed in a different context.

5. Describe a scenario where an object's average speed is a large number, but the magnitude of the average velocity is not a large number.

6. Explain why it is not possible to have an average velocity with a large magnitude but a small average speed.

Problems

In Exercises 7-10, a position function $\vec{r}(t)$ is given. Find $\vec{v}(t)$ and $\vec{a}(t)$.

7. $\vec{r}(t)=\langle 2t+1,5t-2,7\rangle$

8. $\vec{r}(t)=\langle 3t^2-2t+1,-t^2+t+14\rangle$

9. $\vec{r}(t)=\langle\cos t,\sin t\rangle$

10. $\vec{r}(t)=\langle t/10,-\cos t,\sin t\rangle$

In Exercises 11-14, a position function $\vec{r}(t)$ is given. Sketch $\vec{r}(t)$ and $\vec{a}(t)$, then add $\vec{r}(t_0)$ and $\vec{a}(t_0)$ to your sketch, with their initial points at $\vec{r}(t_0)$, for the given value of t_0.

11. $\vec{r}(t) = \langle t, \sin t\rangle$ on $[0, \pi/2]$; $t_0 = \pi/4$

12. $\vec{r}(t) = \langle t^2, \sin t^2\rangle$ on $[0, \pi/2]$; $t_0 = \sqrt{\pi/4}$

13. $\vec{r}(t) = \langle t^2 + t, -t^2 + 2t\rangle$ on $[-2, 2]$; $t_0 = 1$

14. $\vec{r}(t) = \langle \frac{2t+3}{t^2+1}, t^2\rangle$ on $[-1, 1]$; $t_0 = 0$

In Exercises 15-24, a position function $\vec{r}(t)$ of an object is given. Find the speed of the object in terms of t, and find where the speed is minimized/maximized on the indicated interval.

15. $\vec{r}(t) = \langle t^2, t\rangle$ on $[-1, 1]$

16. $\vec{r}(t) = \langle t^2, t^2 - t^3\rangle$ on $[-1, 1]$

17. $\vec{r}(t) = \langle 5\cos t, 5\sin t\rangle$ on $[0, 2\pi]$

18. $\vec{r}(t) = \langle 2\cos t, 5\sin t\rangle$ on $[0, 2\pi]$

19. $\vec{r}(t) = \langle \sec t, \tan t\rangle$ on $[0, \pi/4]$

20. $\vec{r}(t) = \langle t + \cos t, 1 - \sin t\rangle$ on $[0, 2\pi]$

21. $\vec{r}(t) = \langle 12t, 5\cos t, 5\sin t\rangle$ on $[0, 4\pi]$

22. $\vec{r}(t) = \langle t^2 - t, t^2 + t, t\rangle$ on $[0, 1]$

23. $\vec{r}(t) = \left\langle t, t^2, \sqrt{1-t^2}\right\rangle$ on $[-1, 1]$

24. **Projectile Motion**: $\vec{r}(t) = \left\langle (v_0\cos\theta)t, -\frac{1}{2}gt^2 + (v_0\sin\theta)t\right\rangle$ on $\left[0, \frac{2v_0\sin\theta}{g}\right]$.

In Exercises 25-28, position functions $\vec{r}_1(t)$ and $\vec{r}_2(s)$ for two objects are given that follow the same path on the respective intervals.
(a) Show that the positions are the same at the indicated t_0 and s_0 values; ie., show $\vec{r}_1(t_0) = \vec{r}_2(s_0)$.
(b) Find the velocity, speed and acceleration of the two objects at t_0 and s_0, respectively.

25.
$\vec{r}_1(t) = \langle t, t^2\rangle$ on $[0, 1]; t_0 = 1$
$\vec{r}_2(t) = \langle s^2, s^4\rangle$ on $[0, 1]; s_0 = 1$

26.
$\vec{r}_1(t) = \langle 3\cos t, 3\sin t\rangle$ on $[0, 2\pi]; t_0 = \pi/2$
$\vec{r}_2(t) = \langle 3\cos(4s), 3\sin(4s)\rangle$ on $[0, \pi/2]; s_0 = \pi/8$

27.
$\vec{r}_1(t) = \langle 3t, 2t\rangle$ on $[0, 2]; t_0 = 2$
$\vec{r}_2(t) = \langle 6t - 6, 4t - 4\rangle$ on $[1, 2]; s_0 = 2$

28.
$\vec{r}_1(t) = \langle t, \sqrt{t}\rangle$ on $[0, 1]; t_0 = 1$
$\vec{r}_2(t) = \langle \sin t, \sqrt{\sin t}\rangle$ on $[0, \pi/2]; s_0 = \pi/2$

In Exercises 29-32, find the position function of an object given its acceleration and initial velocity and position.

29. $\vec{a}(t) = \langle 2, 3\rangle$; $\vec{v}(0) = \langle 1, 2\rangle$, $\vec{r}(0) = \langle 5, -2\rangle$

30. $\vec{a}(t) = \langle 2, 3\rangle$; $\vec{v}(1) = \langle 1, 2\rangle$, $\vec{r}(1) = \langle 5, -2\rangle$

31. $\vec{a}(t) = \langle \cos t, -\sin t\rangle$; $\vec{v}(0) = \langle 0, 1\rangle$, $\vec{r}(0) = \langle 0, 0\rangle$

32. $\vec{a}(t) = \langle 0, -32\rangle$; $\vec{v}(0) = \langle 10, 50\rangle$, $\vec{r}(0) = \langle 0, 0\rangle$

In Exercises 33-36, find the displacement, distance traveled, average velocity, and average speed of the described object on the given interval.

33. An object with position function $\vec{r}(t) = \langle 2\cos t, 2\sin t, 3t\rangle$, where distances are measured in feet and time is in seconds, on $[0, 2\pi]$.

34. An object with position function $\vec{r}(t)=\langle 5\cos t,-5\sin t\rangle$, where distances are measured in feet and time is in seconds, on $[0,\pi]$.

35. An object with velocity function $\vec{v}(t)=\langle \cos t,\sin t\rangle$, where distances are measured in feet and time is in seconds, on $[0,2\pi]$.

36. An object with velocity function $\vec{r}(t)=\langle 1,2,-1\rangle$, where distances are measured in feet and time is in seconds, on $[0,10]$.

Exercises 37-42 ask you to solve a variety of problems based on the principles of projectile motion.

37. A boy whirls a ball, attached to a 3ft string, above his head in a counter–clockwise circle. The ball makes 2 revolutions per second.
At what *t*-values should the boy release the string so that the ball heads directly for a tree standing 10ft in front of him?

38. David faces Goliath with only a stone in a 3ft sling, which he whirls above his head at 4 revolutions per second. They stand 20ft apart.
(a) At what *t*-values must David release the stone in his sling in order to hit Goliath?
(b) What is the speed at which the stone is traveling when released?
(c) Assume David releases the stone from a height of 6ft and Goliath's forehead is 9ft above the ground. What angle of elevation must David apply to the stone to hit Goliath's head?

39. A hunter aims at a deer which is 40 yards away. Her cross-bow is at a height of 5ft, and she aims for a spot on the deer 4ft above the ground. The crossbow fires her arrows at 300ft/s.
(a) At what angle of elevation should she hold the crossbow to hit her target?
(b) If the deer is moving perpendicularly to her line of sight at a rate of 20mph, by approximately how much should she lead the deer in order to hit in the desired location?

40. A baseball player hits a ball at 100mph, with an intial height of 3ft and an angle of elevation of 20°, at Boston's Fenway Park. The ball flies towards the famed "Green Monster," a wall 37ft high located 310ft from home plate.
(a) Show that as hit, the ball hits the wall.
(b) Show that if the angle of elevation is 21°, the ball clears the Green Monster.

41. A Cessna flies at 1000ft at 150mph and drops a box of supplies to the professor (and his wife) on an island. Ignoring wind resistance, how far horizontally will the supplies travel before they land?

42. A football quarterback throws a pass from a height of 6ft, intending to hit his receiver 20yds away at a height of 5ft.
(a) If the ball is thrown at a rate of 50mph, what angle of elevation is needed to hit his intended target?
(b) If the ball is thrown at with an angle of elevation of 8°, what initial ball speed is needed to hit his target?

11.4: Unit Tangent and Normal Vectors

Terms and Concepts

1. If $\vec{T}(t)$ is a unit tangent vector, what is $\|\vec{T}(t)\|$?

2. If $\vec{N}(t)$ is a unit normal vector, what is $\vec{N}(t)\cdot\vec{r}\,'(t)$?

3. The acceleration vector $\vec{a}(t)$ lies in the plane defined by what two vectors?

4. a_T measures how much the acceleration is affecting the _______ of an object.

Problems

In Exercises 5-8, given $\vec{r}(t)$, find $\vec{T}(t)$ and evaluate it at the indicated value of t.

5. $\vec{r}(t)=\langle 2t^2,t^2-1\rangle,\quad t=1$

6. $\vec{r}(t)=\langle t,\cos t\rangle,\quad t=\pi/4$

7. $\vec{r}(t)=\langle \cos^3 t,\sin^3 t\rangle,\quad t=\pi/4$

8. $\vec{r}(t)=\langle \cos t,\sin t\rangle,\quad t=\pi$

In Exercises 9-12, find the equation of the line tangent to the curve at the indicated *t*-value using the unit tangent vector. Note: these are the same problems as in Exercises 5-8.

9. $\vec{r}(t)=\langle 2t^2,t^2-1\rangle,\quad t=1$

10. $\vec{r}(t)=\langle t,\cos t\rangle,\quad t=\pi/4$

11. $\vec{r}(t) = \langle \cos^3 t, \sin^3 t \rangle, \quad t = \pi/4$

12. $\vec{r}(t) = \langle \cos t, \sin t \rangle, \quad t = \pi$

In Exercises 13-16, find $\vec{N}(t)$ using Definition 75. Confirm the result using the Theorem 97.

13. $\vec{r}(t) = \langle 3\cos t, 3\sin t \rangle$

14. $\vec{r}(t) = \langle t, t^2 \rangle$

15. $\vec{r}(t) = \langle \cos t, 2\sin t \rangle$

16. $\vec{r}(t) = \langle e^t, e^{-t} \rangle$

In Exercises 17-20, a position function $\vec{r}(t)$ is given along with its unit tangent vector $\vec{T}(t)$ evaluated at $t = a$, for some value of a.
(a) Confirm that $\vec{T}(a)$ is as stated.
(b) Using a graph of $\vec{r}(t)$ and Theorem 97, find $\vec{N}(a)$.

17. $\vec{r}(t) = \langle 3\cos t, 5\sin t \rangle; \quad \vec{T}(\pi/4) = \left\langle -\frac{3}{\sqrt{34}}, \frac{5}{\sqrt{34}} \right\rangle$

18. $\vec{r}(t) = \left\langle t, \frac{1}{t^2+1} \right\rangle; \quad \vec{T}(1) = \left\langle \frac{2}{\sqrt{5}}, -\frac{1}{\sqrt{5}} \right\rangle$

19. $\vec{r}(t) = (1 + 2\sin t)(\cos t, \sin t); \quad \vec{T}(0) = \left\langle \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}} \right\rangle$

20. $\vec{r}(t) = \langle \cos^3 t, \sin^3 t \rangle; \quad \vec{T}(\pi/4) = \left\langle -\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right\rangle$

In Exercises 21-24, find $\vec{N}(t)$.

21. $\vec{r}(t) = \langle 4t, 2\sin t, 2\cos t \rangle$

22. $\vec{r}(t) = \langle 5\cos t, 3\sin t, 4\sin t \rangle$

23. $\vec{r}(t) = \langle a\cos t, a\sin t, bt \rangle$

24. $\vec{r}(t) = \langle \cos(at), \sin(at), t \rangle$

In Exercises 25-30, find a_T and a_N given $\vec{r}(t)$. Sketch $\vec{r}(t)$ on the indicated interval, and comment on the relative sizes of a_T and a_N at the indicated *t* values.

25. $\vec{r}(t) = \langle t, t^2 \rangle$ on $[-1, 1]$; consider $t = 0$ and $t = 1$.

26. $\vec{r}(t) = \langle t, 1/t \rangle$ on $(0, 4$; consider $t = 1$ and $t = 2$.

27. $\vec{r}(t) = \langle 2\cos t, 2\sin t \rangle$ on $[0, 2\pi]$; consider $t = 1$ and $t = \pi/2$.

28. $\vec{r}(t) = \langle \cos(t^2), \sin(t^2) \rangle$ on $(0, 2\pi]$; consider $t = \sqrt{\pi/2}$ and $t = \sqrt{\pi}$.

29. $\vec{r}(t) = \langle a\cos t, a\sin t, bt \rangle$ on $[0, 2\pi]$; where $a, b > 0$; consider $t = 0$ and $t = \pi/2$.

30. $\vec{r}(t) = \langle 5\cos t, 4\sin t, 3\sin t \rangle$ on $[0, 2\pi]$; consider $t = 0$ and $t = \pi/2$.

11.5: The Arc Length Parameter and Curvature

Terms and Concepts

1. It is common to describe position in terms of both ______ and/or _______.

2. A measure of the "curviness" of a curve is ________.

3. Give two shapes with constant curvature.

4. Describe in your own words what an "osculating circle" is.

5. Complete the identity: $\vec{T}'(s) =$_________$\vec{N}(s)$.

6. Given a position function $\vec{r}(t)$, how are a_T and a_N affected by the curvature?

Problems

In Exercises 7-10, a position function $\vec{r}(t)$ is given, where \(t=0\) corresponds to the initial position. Find the arc length parameter s, and rewrite $\vec{r}(t)$ in terms of s; that is, find $\vec{r}(s)$.

7. $\vec{r}(t)=\langle 2t,t,-2t\rangle$

8. $\vec{r}(t)=\langle 7\cos t,7\sin t\rangle$

9. $\vec{r}(t)=\langle 3\cos t,3\sin t,2t\rangle$

10. $\vec{r}(t)=\langle 5\cos t,13\sin t,12\cos t\rangle$

In Exercises 11-22, a curve *C* is described along with 2 points on *C*.
(a) Using a sketch, determine at which of these points the curvature is greater.
(b) Find the curvature κ of *C*, and evaluate κ at each of the 2 given points.

11. C is defined by $y=x^3-x$; points given at $x=0$ and $x=1/2$.

12. C is defined by $y=\frac{1}{x^2+1}$; points given at $x=0$ and $x=2$.

13. C is defined by $y=\cos x$; points given at $x=0$ and $x=\pi/2$.

14. C is defined by $y=\sqrt{1-x^2}$ on $(-1,1)$; points given at $x=0$ and $x=1/2$.

15. C is defined by $\vec{r}(t)=\langle\cos t,\sin(2t)\rangle$; points given at $t=0$ and $t=\pi/4$.

16. C is defined by $\vec{r}(t)=\langle\cos^2(t),\sin t\cos t\rangle$; points given at $t=0$ and $t=\pi/3$.

17. C is defined by $\vec{r}(t)=\langle t^2-1,t^3-t\rangle$; points given at $t=0$ and $t=\pi/6$.

18. C is defined by $\vec{r}(t)=\langle\tan t,\sec t\rangle$; points given at $t=0$ and $t=\pi/6$.

19. C is defined by $\vec{r}(t)=\langle 4t+2,3t-1,2t+5\rangle$; points given at $t=0$ and $t=1$.

20. C is defined by $\vec{r}(t)=\langle t^3-t,t^3-4,t^2-1\rangle$; points given at $t=0$ and $t=1$.

21. C is defined by $\vec{r}(t)=\langle 3\cos t,3\sin t,2t\rangle$; points given at $t=0$ and $t=\pi/2$.

22. C is defined by $\vec{r}(t)=\langle 5\cos t,13\sin t,12\cos t\rangle$; points given at $t=0$ and $t=\pi/2$.

In Exercises 23-26, find the value of *x* or *t* where curvature is maximized.

23. $y=\frac{1}{6}x^3$

24. $y=\sin x$

25. $\vec{r}(t)=\langle t^2+2t,3t-t^2\rangle$

26. $\vec{r}(t)=\langle t,4/t,3/t\rangle$

In Exercises 27-30, find the radius of curvature at the indicated value.

27. $y=\tan x,\ \text{at } x=\pi/4$

28. $y=x^2+x-3,\ \text{at } x=\pi/4$

29. $\vec{r}(t)=\langle\cos t,\sin(3t)\rangle,\ \text{at } t=0$

30. $\vec{r}(t)=\langle 5\cos(3t),t\rangle,\ \text{at } t=0$

In Exercises 31-34, find the equation of the osculating circle to the curve at the indicated *t*-value.

31. $\vec{r}(t)=\langle t,t^2\rangle,\ \text{at } t=0.$

32. $\vec{r}(t)=\langle 3\cos t,\sin t\rangle,\ \text{at } t=0.$

33. $\vec{r}(t)=\langle 3\cos t,\sin t\rangle,\ \text{at } t=\pi/2.$

34. $\vec{r}(t)=\langle t^2-t,t^2+t\rangle,\ \text{at } t=0.$

11.E: Applications of Vector Valued Functions (Exercises) is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

12: Functions of Several Variables

A function of the form $y = f(x)$ is a function of a single variable; given a value of x, we can find a value y. Even the vector--valued functions of Chapter 11 are single--variable functions; the input is a single variable though the output is a vector. There are many situations where a desired quantity is a function of two or more variables. For instance, wind chill is measured by knowing the temperature and wind speed; the volume of a gas can be computed knowing the pressure and temperature of the gas; to compute a baseball player's batting average, one needs to know the number of hits and the number of at--bats. This chapter studies **multivariable** functions, that is, functions with more than one input.

12.1: Introduction to Multivariable Functions
12.2: Limits and Continuity of Multivariable Functions
12.3: Partial Derivatives
12.4: Differentiability and the Total Differential
12.5: The Multivariable Chain Rule
12.6: Directional Derivatives
12.7: Tangent Lines, Normal Lines, and Tangent Planes
12.8: Extreme Values
12.E: Applications of Functions of Several Variables (Exercises)

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 12: Functions of Several Variables is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

12.1: Introduction to Multivariable Functions

Definition 77 Function of Two Variables

Let D be a subset of $\mathbb{R}^2$. A **function** f **of two variables** is a rule that assigns each pair (x, y) in D a value $z = f(x, y)$ in $\mathbb{R}$. D is the **domain** of f; the set of all outputs of f is the **range**.

Example 12.1.1: Understanding a function of two variables

Let $z = f(x, y) = x^2 - y$. Evaluate $f(1, 2)$, $f(2, 1)$, and $f(-2, 4)$; find the domain and range of f.

Solution

Using the definition $f(x, y) = x^2 - y$, we have:

$$\begin{aligned} f(1,2) &= 1^2 - 2 = -1 \\ f(2,1) &= 2^2 - 1 = 3 \\ f(-2,4) &= (-2)^2 - 4 = 0 \end{aligned}$$

The domain is not specified, so we take it to be all possible pairs in $\mathbb{R}^2$ for which f is defined. In this example, f is defined for *all* pairs (x, y), so the domain D of f is $\mathbb{R}^2$.

The output of f can be made as large or small as possible; any real number r can be the output. (In fact, given any real number r, $f(0, -r) = r$.) So the range R of f is $\mathbb{R}$.

Example 12.1.2: Understanding a function of two variables

Let

$$f(x,y) = \sqrt{1 - \frac{x^2}{9} - \frac{y^2}{4}}.$$

Find the domain and range of f.

Solution

The domain is all pairs (x, y) allowable as input in f. Because of the square-root, we need (x, y) such that $0 \le 1 - \frac{x^2}{9} - \frac{y^2}{4}$:

$$\begin{aligned} 0 &\le 1 - \frac{x^2}{9} - \frac{y^2}{4} \\ \frac{x^2}{9} + \frac{y^2}{4} &\le 1 \end{aligned}$$

The above equation describes the interior of an ellipse as shown in Figure 12.1.1. We can represent the domain D graphically with the figure; in set notation, we can write $D = \{(x, y) | \frac{x^2}{9} + \frac{y^2}{4} \le 1\}$.

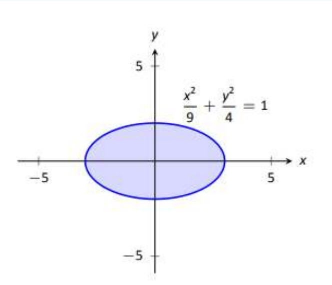


Figure 12.1.1: Illustrating the domain of $f(x, y)$ in Example 12.1.2

The range is the set of all possible output values. The square-root ensures that all output is ≥ 0. Since the x and y terms are squared, then subtracted, inside the square-root, the largest output value comes at $x = 0$, $y = 0$: $f(0, 0) = 1$. Thus the range R is the interval $[0, 1]$.

Graphing Functions of Two Variables

The **graph** of a function f of two variables is the set of all points $\big(x,y,f(x,y)\big)$ where (x,y) is in the domain of f. This creates a **surface** in space.

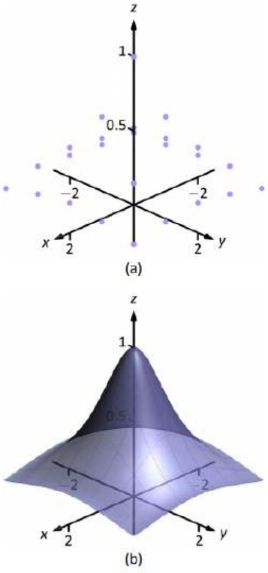


Figure $12.1.2$: Graphing a function of two variables.

One can begin sketching a graph by plotting points, but this has limitations. Consider Figure $12.1.2a$ where 25 points have been plotted of $f(x,y)=\frac{1}{x^2+y^2+1}$. More points have been plotted than one would reasonably want to do by hand, yet it is not clear at all what the graph of the function looks like. Technology allows us to plot lots of points, connect adjacent points with lines and add shading to create a graph like Figure $12.1.2b$ which does a far better job of illustrating the behavior of f.

While technology is readily available to help us graph functions of two variables, there is still a paper-and-pencil approach that is useful to understand and master as it, combined with high-quality graphics, gives one great insight into the behavior of a function. This technique is known as sketching **level curves**.

Level Curves

It may be surprising to find that the problem of representing a three dimensional surface on paper is familiar to most people (they just don't realize it). Topographical maps, like the one shown in Figure $12.1.3$, represent the surface of Earth by indicating points with the same elevation with **contour lines**. The elevations marked are equally spaced; in this example, each thin line indicates an elevation change in 50 ft increments and each thick line indicates a change of 200 ft. When lines are drawn close together, elevation changes rapidly (as one does not have to travel far to rise 50 ft). When lines are far apart, such as near "Aspen Campground," elevation changes more gradually as one has to walk farther to rise 50 ft.

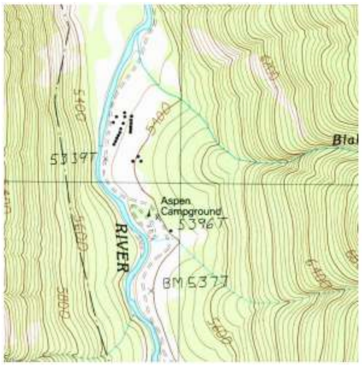


Figure $12.1.3$: A topographical map displays elevation by drawing contour lines, along with the elevation is constant.

Given a function $z=f(x,y)$, we can draw a "topographical map" of f by drawing **level curves** (or, contour lines). A level curve at $z=c$ is a curve in the x-y plane such that for all points (x,y) on the curve, $f(x,y)=c$.

When drawing level curves, it is important that the c values are spaced equally apart as that gives the best insight to how quickly the "elevation" is changing. Examples will help one understand this concept.

Example $12.1.3$: Drawing Level Curves

Let $f(x,y)=\sqrt{1-\frac{x^2}{9}-\frac{y^2}{4}}$. Find the level curves of f for $c=0$, 0.2, 0.4, 0.6, 0.8 and 1.

Solution

Consider first $c=0$. The level curve for $c=0$ is the set of all points (x,y) such that $0=\sqrt{1-\frac{x^2}{9}-\frac{y^2}{4}}$. Squaring both sides gives us

$$\frac{x^2}{9}+\frac{y^2}{4}=1, \tag{12.1.1}$$

an ellipse centered at $(0,0)$ with horizontal major axis of length 6 and minor axis of length 4. Thus for any point (x,y) on this curve, $f(x,y)=0$.

Now consider the level curve for $c=0.2$

$$\begin{aligned}0.2&=\sqrt{1-\frac{x^2}{9}-\frac{y^2}{4}}\\0.04&=1-\frac{x^2}{9}-\frac{y^2}{4}\\\frac{x^2}{9}+\frac{y^2}{4}&=0.96\\\frac{x^2}{8.64}+\frac{y^2}{3.84}&=1.\end{aligned}$$

This is also an ellipse, where $a=\sqrt{8.64}\approx 2.94$ and $b=\sqrt{3.84}\approx 1.96$.

In general, for $z=c$, the level curve is:

$$\begin{aligned}c&=\sqrt{1-\frac{x^2}{9}-\frac{y^2}{4}}\\c^2&=1-\frac{x^2}{9}-\frac{y^2}{4}\\\frac{x^2}{9}+\frac{y^2}{4}&=1-c^2\\\frac{x^2}{9(1-c^2)}+\frac{y^2}{4(1-c^2)}&=1,\end{aligned}$$

ellipses that are decreasing in size as c increases. A special case is when $c=1$; there the ellipse is just the point $(0,0)$.

The level curves are shown in Figure 12.4(a). Note how the level curves for $c=0$ and $c=0.2$ are very, very close together: this indicates that f is growing rapidly along those curves.

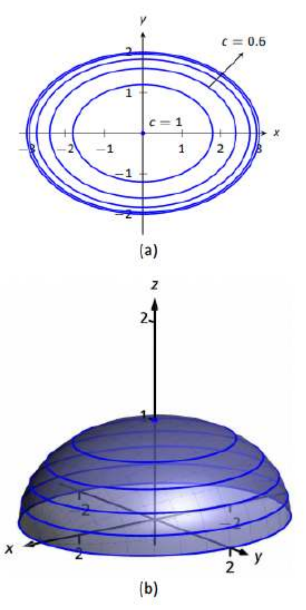


Figure $12.1.4$: Graphing the level curves in Example 12.1.4

In Figure $12.1.4b$, the curves are drawn on a graph of f in space. Note how the elevations are evenly spaced. Near the level curves of $c = 0$ and $c = 0.2$ we can see that f indeed is growing quickly.

Example $12.1.4$: Analyzing Level Curves

Let $f(x, y) = \frac{x+y}{x^2+y^2+1}$. Find the level curves for $z = c$.

Solution

We begin by setting $f(x, y) = c$ for an arbitrary c and seeing if algebraic manipulation of the equation reveals anything significant.

$$\begin{aligned} \frac{x+y}{x^2+y^2+1} &= c \\ x+y &= c(x^2+y^2+1). \end{aligned}$$

We recognize this as a circle, though the center and radius are not yet clear. By completing the square, we can obtain:

$$\left(x - \frac{1}{2c}\right)^2 + \left(y - \frac{1}{2c}\right)^2 = \frac{1}{2c^2} - 1, \tag{12.1.2}$$

a circle centered at $\big(1/(2c), 1/(2c)\big)$ with radius $\sqrt{1/(2c^2) - 1}$, where $|c| < 1/\sqrt{2}$. The level curves for $c = \pm 0.2, \ \pm 0.4$ and ± 0.6 are sketched in Figure $12.1.5a$. To help illustrate "elevation," we use thicker lines for c values near 0, and dashed lines indicate where $c < 0$.

There is one special level curve, when $c = 0$. The level curve in this situation is $x + y = 0$, the line $y = -x$.

In Figure $12.1.5b$ we see a graph of the surface. Note how the y-axis is pointing away from the viewer to more closely resemble the orientation of the level curves in (a).

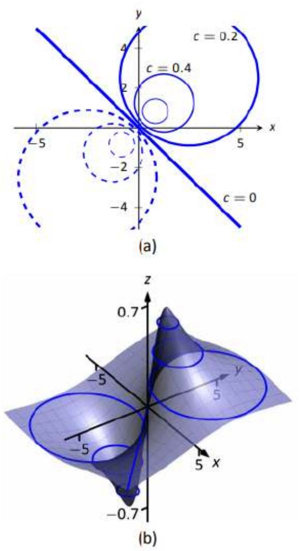


Figure $12.1.5$: Graphing the level curves in Example 12.1.4

Seeing the level curves helps us understand the graph. For instance, the graph does not make it clear that one can "walk" along the line $y=-x$ without elevation change, though the level curve does.

Functions of Three Variables

We extend our study of multivariable functions to functions of three variables. (One can make a function of as many variables as one likes; we limit our study to three variables.)

Definition 78 Function of Three Variables

Let D be a subset of $\mathbb{R}^3$. A **function** f **of three variables** is a rule that assigns each triple (x,y,z) in D a value $w=f(x,y,z)$ in $\mathbb{R}$. D is the **domain** of f; the set of all outputs of f is the **range**.

Note how this definition closely resembles that of Definition 77.

Example $12.1.5$: Understanding a function of three variables

Let

$$f(x,y,z)=\frac{x^2+z+3\sin y}{x+2y-z}.$$

Evaluate f at the point $(3,0,2)$ and find the domain and range of f.

Solution

$$f(3,0,2)=\frac{3^2+2+3\sin 0}{3+2(0)-2}=11.$$

As the domain of f is not specified, we take it to be the set of all triples (x,y,z) for which $f(x,y,z)$ is defined. As we cannot divide by 0, we find the domain D is

$$D=\{(x,y,z)\mid x+2y-z\neq 0\}.$$

We recognize that the set of all points in $\mathbb{R}^3$ that *are not* in D form a plane in space that passes through the origin (with normal vector $\langle 1,2,-1\rangle$).

We determine the range R is $\mathbb{R}$; that is, all real numbers are possible outputs of f. There is no set way of establishing this. Rather, to get numbers near 0 we can let $y=0$ and choose $z\approx -x^2$. To get numbers of arbitrarily large magnitude, we can let $z\approx x+2y$.

Level Surfaces

It is very difficult to produce a meaningful graph of a function of three variables. A function of *one* variable is a *curve* drawn in *2* dimensions; a function of *two* variables is a *surface* drawn in *3* dimensions; a function of *three* variables is a *hypersurface* drawn in *4* dimensions.

There are a few techniques one can employ to try to "picture" a graph of three variables. One is an analogue of level curves: **level surfaces**. Given $w = f(x, y, z)$, the level surface at $w = c$ is the surface in space formed by all points (x, y, z) where $f(x, y, z) = c$.

Example $12.1.6$: Finding level surfaces

If a point source S is radiating energy, the intensity I at a given point P in space is inversely proportional to the square of the distance between S and P. That is, when $S = (0, 0, 0)$, $I(x, y, z) = \frac{k}{x^2+y^2+z^2}$ for some constant k.

Let $k = 1$; find the level surfaces of I.

Solution

We can (mostly) answer this question using "common sense." If energy (say, in the form of light) is emanating from the origin, its intensity will be the same at all points equidistant from the origin. That is, at any point on the surface of a sphere centered at the origin, the intensity should be the same. Therefore, the level surfaces are spheres.

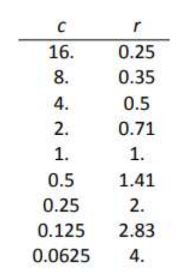
We now find this mathematically. The level surface at $I = c$ is defined by

$$c = \frac{1}{x^2 + y^2 + z^2}. \tag{12.1.3}$$

A small amount of algebra reveals

$$x^2 + y^2 + z^2 = \frac{1}{c}. \tag{12.1.4}$$

Given an intensity c, the level surface $I = c$ is a sphere of radius $1/\sqrt{c}$, centered at the origin.



c	r
16.	0.25
8.	0.35
4.	0.5
2.	0.71
1.	1.
0.5	1.41
0.25	2.
0.125	2.83
0.0625	4.

Figure $12.1.6$: A table of c values and the corresponding radius r of the spheres of constant value in Example 12.1.6

Figure $12.1.6$ gives a table of the radii of the spheres for given c values. Normally one would use equally spaced c values, but these values have been chosen purposefully. At a distance of 0.25 from the point source, the intensity is 16; to move to a point of half that intensity, one just moves out 0.1 to 0.35 - not much at all. To again halve the intensity, one moves 0.15, a little more than before.

Note how each time the intensity if halved, the distance required to move away grows. We conclude that the closer one is to the source, the more rapidly the intensity changes.

In the next section we apply the concepts of limits to functions of two or more variables.

This page titled 12.1: Introduction to Multivariable Functions is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

12.2: Limits and Continuity of Multivariable Functions

We continue with the pattern we have established in this text: after defining a new kind of function, we apply calculus ideas to it. The previous section defined functions of two and three variables; this section investigates what it means for these functions to be "continuous."

We begin with a series of definitions. We are used to "open intervals" such as $(1,3)$, which represents the set of all x such that $1 < x < 3$, and "closed intervals" such as $[1,3]$, which represents the set of all x such that $1 \leq x \leq 3$. We need analogous definitions for open and closed sets in the x-y plane.

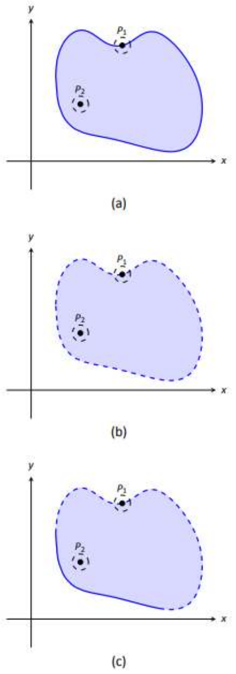
Definition 79 Open Disk, Boundary and Interior Points, Open and Closed Sets, Bounded Sets

An **open disk** B in $\mathbb{R}^2$ centered at (x_0, y_0) with radius r is the set of all points (x,y) such that $\sqrt{(x-x_0)^2+(y-y_0)^2} < r$.

Let S be a set of points in $\mathbb{R}^2$. A point P in $\mathbb{R}^2$ is a **boundary point** of S if all open disks centered at P contain both points in S and points not in S.

- A point P in S is an **interior point** of S if there is an open disk centered at P that contains only points in S.
- A set S is **open** if every point in S is an interior point.
- A set S is **closed** if it contains all of its boundary points.
- A set S is **bounded** if there is an $M > 0$ such that the open disk, centered at the origin with radius M, contains S. A set that is not bounded is **unbounded**.

Figure 12.7 shows several sets in the x-y plane. In each set, point P_1 lies on the boundary of the set as all open disks centered there contain both points in, and not in, the set. In contrast, point P_2 is an interior point for there is an open disk centered there that lies entirely within the set.



(a)

(b)

(c)

Figure 12.7: Illustrating open and closed sets in the x-y plane.

The set depicted in Figure 12.7(a) is a closed set as it contains all of its boundary points. The set in (b) is open, for all of its points are interior points (or, equivalently, it does not contain any of its boundary points). The set in (c) is neither open nor closed as it contains some of its boundary points.

Example 12.2.1: Determining open/closed, bounded/unbounded

Determine if the domain of the function $f(x,y)=\sqrt{1-\frac{x^2}{9}-\frac{y^2}{4}}$ is open, closed, or neither, and if it is bounded.

Solution

This domain of this function was found in Example 12.1.1 to be $D=\{(x,y)\mid \frac{x^2}{9}+\frac{y^2}{4}\leq 1\}$, the region *bounded* by the ellipse $\frac{x^2}{9}+\frac{y^2}{4}=1$. Since the region includes the boundary (indicated by the use of "$\leq$"), the set contains all of its boundary points and hence is closed. The region is bounded as a disk of radius 4, centered at the origin, contains D.

Example 12.2.2: Determining open/closed, bounded/unbounded

Determine if the domain of $f(x,y)=\frac{1}{x-y}$ is open, closed, or neither.

Solution

As we cannot divide by 0, we find the domain to be $D=\{(x,y)\mid x-y\neq 0\}$. In other words, the domain is the set of all points (x,y) *not* on the line $y=x$.

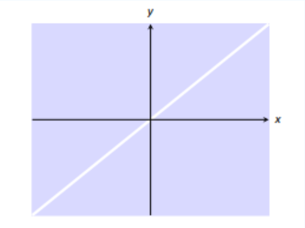


Figure 12.8: Sketching the domain of the function in Example 12.2.2

The domain is sketched in Figure 12.8. Note how we can draw an open disk around any point in the domain that lies entirely inside the domain, and also note how the only boundary points of the domain are the points on the line $y=x$. We conclude the domain is an open set. The set is unbounded.

Limits

Recall a pseudo--definition of the limit of a function of one variable: "$\lim\limits_{x\to c} f(x)=L$" means that if x is "really close" to c, then $f(x)$ is "really close" to L. A similar pseudo--definition holds for functions of two variables. We'll say that

$$''\lim_{(x,y)\to(x_0,y_0)} f(x,y)=L'' \tag{12.2.1}$$

means "if the point (x,y) is really close to the point (x_0,y_0), then $f(x,y)$ is really close to L." The formal definition is given below.

Definition 80 Limit of a Function of Two Variables

Let S be an open set containing (x_0,y_0), and let f be a function of two variables defined on S, except possibly at (x_0,y_0). The **limit** of $f(x,y)$ as (x,y) approaches (x_0,y_0) is L, denoted

$$\lim_{(x,y)\to(x_0,y_0)} f(x,y)=L, \tag{12.2.2}$$

means that given any $\epsilon>0$, there exists $\delta>0$ such that for all $(x,y)\neq(x_0,y_0)$, if (x,y) is in the open disk centered at (x_0,y_0) with radius δ, then $|f(x,y)-L|<\epsilon$.

The concept behind Definition 80 is sketched in Figure 12.9. Given $\epsilon>0$, find $\delta>0$ such that if (x,y) is any point in the open disk centered at (x_0,y_0) in the x-y plane with radius δ, then $f(x,y)$ should be within ϵ of L.

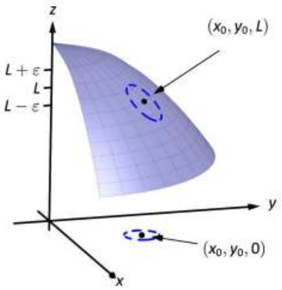


Figure 12.9: Illustrating the definition of a limit. The open disk in the x-y plane has radius δ. Let (x,y) be any point in this disk; $f(x,y)$ is within ϵ of L.

Computing limits using this definition is rather cumbersome. The following theorem allows us to evaluate limits much more easily.

THEOREM 101 Basic Limit Properties of Functions of Two Variables

Let b, x_0, y_0, L and K be real numbers, let n be a positive integer, and let f and g be functions with the following limits:

$$\lim_{(x,y)\to(x_0,y_0)} f(x,y) = L \quad \text{and} \quad \lim_{(x,y)\to(x_0,y_0)} g(x,y) = K. \tag{12.2.3}$$

The following limits hold.

1. Constants: $\lim_{(x,y)\to(x_0,y_0)} b = b$
2. Identity : $\lim_{(x,y)\to(x_0,y_0)} x = x_0;\qquad \lim_{(x,y)\to(x_0,y_0)} y = y_0$
3. Sums/Differences: $\lim_{(x,y)\to(x_0,y_0)} \big(f(x,y) \pm g(x,y)\big) = L \pm K$
4. Scalar Multiples: $\lim_{(x,y)\to(x_0,y_0)} b\cdot f(x,y) = bL$
5. Products: $\lim_{(x,y)\to(x_0,y_0)} f(x,y)\cdot g(x,y) = LK$
6. Quotients: $\lim_{(x,y)\to(x_0,y_0)} f(x,y)/g(x,y) = L/K,\ (K \neq 0)$
7. Powers: $\lim_{(x,y)\to(x_0,y_0)} f(x,y)^n = L^n$

This theorem, combined with Theorems 2 and 3 of Section 1.3, allows us to evaluate many limits.

Example 12.2.3: Evaluating a limit

Evaluate the following limits:

$$1.\ \lim_{(x,y)\to(1,\pi)} \frac{y}{x} + \cos(xy) \qquad 2.\ \lim_{(x,y)\to(0,0)} \frac{3xy}{x^2+y^2} \tag{12.2.4}$$

Solution

1. The aforementioned theorems allow us to simply evaluate $y/x + \cos(xy)$ when $x = 1$ and $y = \pi$. If an indeterminate form is returned, we must do more work to evaluate the limit; otherwise, the result is the limit. Therefore

$$\begin{aligned}\lim_{(x,y)\to(1,\pi)} \frac{y}{x} + \cos(xy) &= \frac{\pi}{1} + \cos\pi \\ &= \pi - 1.\end{aligned}$$

2. We attempt to evaluate the limit by substituting 0 in for x and y, but the result is the indeterminate form "$0/0$." To evaluate this limit, we must "do more work," but we have not yet learned what "kind" of work to do. Therefore we cannot yet evaluate this limit.

When dealing with functions of a single variable we also considered one--sided limits and stated

$$\lim_{x\to c} f(x) = L \quad \text{if, and only if,} \quad \lim_{x\to c^+} f(x) = L \quad \textbf{and} \quad \lim_{x\to c^-} f(x) = L. \tag{12.2.5}$$

That is, the limit is L if and only if $f(x)$ approaches L when x approaches c from **either** direction, the left or the right.

In the plane, there are infinite directions from which (x,y) might approach (x_0,y_0). In fact, we do not have to restrict ourselves to approaching (x_0,y_0) from a particular direction, but rather we can approach that point along a path that is not a straight line. It is possible to arrive at different limiting values by approaching (x_0,y_0) along different paths. If this happens, we say that $\lim\limits_{(x,y)\to(x_0,y_0)} f(x,y)$ does not exist (this is analogous to the left and right hand limits of single variable functions not being equal).

Our theorems tell us that we can evaluate most limits quite simply, without worrying about paths. When indeterminate forms arise, the limit may or may not exist. If it does exist, it can be difficult to prove this as we need to show the same limiting value is obtained regardless of the path chosen. The case where the limit does not exist is often easier to deal with, for we can often pick two paths along which the limit is different.

Example 12.2.4: Showing limits do not exist

1. Show $\lim\limits_{(x,y)\to(0,0)} \frac{3xy}{x^2+y^2}$ does not exist by finding the limits along the lines $y=mx$.
2. Show $\lim\limits_{(x,y)\to(0,0)} \frac{\sin(xy)}{x+y}$ does not exist by finding the limit along the path $y=-\sin x$.

Solution

1. Evaluating $\lim\limits_{(x,y)\to(0,0)} \frac{3xy}{x^2+y^2}$ along the lines $y=mx$ means replace all y's with mx and evaluating the resulting limit:

$$\begin{aligned}\lim_{(x,mx)\to(0,0)} \frac{3x(mx)}{x^2+(mx)^2} &= \lim_{x\to 0}\frac{3mx^2}{x^2(m^2+1)}\\ &= \lim_{x\to 0}\frac{3m}{m^2+1}\\ &= \frac{3m}{m^2+1}.\end{aligned}$$

   While the limit exists for each choice of m, we get a *different* limit for each choice of m. That is, along different lines we get differing limiting values, meaning *the* limit does not exist.

2. Let $f(x,y)=\frac{\sin(xy)}{x+y}$. We are to show that $\lim\limits_{(x,y)\to(0,0)} f(x,y)$ does not exist by finding the limit along the path $y=-\sin x$. First, however, consider the limits found along the lines $y=mx$ as done above.

$$\begin{aligned}\lim_{(x,mx)\to(0,0)} \frac{\sin\big(x(mx)\big)}{x+mx} &= \lim_{x\to 0}\frac{\sin(mx^2)}{x(m+1)}\\ &= \lim_{x\to 0}\frac{\sin(mx^2)}{x}\cdot\frac{1}{m+1}.\end{aligned}$$

   By applying L'H\^opital's Rule, we can show this limit is 0 *except* when $m=-1$, that is, along the line $y=-x$. This line is not in the domain of f, so we have found the following fact: along every line $y=mx$ in the domain of f, $\lim\limits_{(x,y)\to(0,0)} f(x,y)$ $=0$.

   Now consider the limit along the path $y=-\sin x$:

$$\lim_{(x,-\sin x)\to(0,0)} \frac{\sin\big(-x\sin x\big)}{x-\sin x} = \lim_{x\to 0}\frac{\sin\big(-x\sin x\big)}{x-\sin x} \tag{12.2.6}$$

   Now apply L'H\^opital's Rule twice:

$$\begin{aligned}&= \lim_{x\to 0}\frac{\cos\big(-x\sin x\big)(-\sin x - x\cos x)}{1-\cos x} \qquad (\text{"} = 0/0\text{”})\\ &= \lim_{x\to 0}\frac{-\sin\big(-x\sin x\big)(-\sin x - x\cos x)^2+\cos\big(-x\sin x\big)(-2\cos x + x\sin x)}{\sin x}\\ &= \text{"}2/0\text{”} \Rightarrow \text{the limit does not exist.}\end{aligned}$$

Step back and consider what we have just discovered. Along any line $y = mx$ in the domain of the $f(x, y)$, the limit is 0. However, along the path $y = -\sin x$, which lies in the domain of $f(x, y)$ for all $x \neq 0$, the limit does not exist. Since the limit is not the same along every path to $(0, 0)$, we say $\lim\limits_{(x,y)\to(0,0)} \frac{\sin(xy)}{x+y}$ does not exist.

Example 12.2.5: Finding a limit

Let $f(x, y) = \frac{5x^2y^2}{x^2+y^2}$. Find $\lim\limits_{(x,y)\to(0,0)} f(x, y)$.

Solution

It is relatively easy to show that along any line $y = mx$, the limit is 0. This is not enough to prove that the limit exists, as demonstrated in the previous example, but it tells us that if the limit does exist then it must be 0.

To prove the limit is 0, we apply Definition 80. Let $\epsilon > 0$ be given. We want to find $\delta > 0$ such that if $\sqrt{(x-0)^2 + (y-0)^2} < \delta$, then $|f(x, y) - 0| < \epsilon$.

Set $\delta < \sqrt{\epsilon/5}$. Note that $\left|\frac{5y^2}{x^2+y^2}\right| < 5$ for all $(x, y) \neq (0, 0)$, and that if $\sqrt{x^2+y^2} < \delta$, then $x^2 < \delta^2$.

Let $\sqrt{(x-0)^2 + (y-0)^2} = \sqrt{x^2+y^2} < \delta$. Consider $|f(x, y) - 0|$:

$$\begin{aligned} |f(x,y) - 0| &= \left|\frac{5x^2y^2}{x^2+y^2} - 0\right| \\ &= \left|x^2 \cdot \frac{5y^2}{x^2+y^2}\right| \\ &< \delta^2 \cdot 5 \\ &< \frac{\epsilon}{5} \cdot 5 \\ &= \epsilon. \end{aligned}$$

Thus if $\sqrt{(x-0)^2 + (y-0)^2} < \delta$ then $|f(x, y) - 0| < \epsilon$, which is what we wanted to show. Thus $\lim\limits_{(x,y)\to(0,0)} \frac{5x^2y^2}{x^2+y^2} = 0$.

Continuity

Definition 3 defines what it means for a function of one variable to be continuous. In brief, it meant that the graph of the function did not have breaks, holes, jumps, etc. We define continuity for functions of two variables in a similar way as we did for functions of one variable.

Definition 81 Continuous

Let a function $f(x, y)$ be defined on an open disk B containing the point (x_0, y_0).

1. f is **continuous** at (x_0, y_0) if $\lim\limits_{(x,y)\to(x_0,y_0)} f(x, y) = f(x_0, y_0)$.
2. f is **continuous on** B if f is continuous at all points in B. If f is continuous at all points in $\mathbb{R}^2$, we say that f is **continuous everywhere**.

Example 12.2.6: Continuity of a function of two variables

Let $f(x, y) = \begin{cases} \frac{\cos y \sin x}{x} & x \neq 0 \\ \cos y & x = 0 \end{cases}$. Is f continuous at $(0, 0)$? Is f continuous everywhere?

Solution

To determine if f is continuous at $(0, 0)$, we need to compare $\lim\limits_{(x,y)\to(0,0)} f(x, y)$ to $f(0, 0)$.

Applying the definition of f, we see that $f(0, 0) = \cos 0 = 1$.

We now consider the limit $\lim\limits_{(x,y)\to(0,0)} f(x, y)$. Substituting 0 for x and y in $(\cos y \sin x)/x$ returns the indeterminate form "0/0", so we need to do more work to evaluate this limit.

Consider two related limits: $\lim\limits_{(x,y)\to(0,0)} \cos y$ and $\lim\limits_{(x,y)\to(0,0)} \frac{\sin x}{x}$. The first limit does not contain x, and since $\cos y$ is continuous,

$$\lim_{(x,y)\to(0,0)} \cos y = \lim_{y\to 0} \cos y = \cos 0 = 1. \tag{12.2.7}$$

The second limit does not contain y. By Theorem 5 we can say

$$\lim_{(x,y)\to(0,0)} \frac{\sin x}{x} = \lim_{x\to 0} \frac{\sin x}{x} = 1. \tag{12.2.8}$$

Finally, Theorem 101 of this section states that we can combine these two limits as follows:

$$\begin{aligned}\lim_{(x,y)\to(0,0)} \frac{\cos y \sin x}{x} &= \lim_{(x,y)\to(0,0)} (\cos y)\left(\frac{\sin x}{x}\right)\\ &= \left(\lim_{(x,y)\to(0,0)} \cos y\right)\left(\lim_{(x,y)\to(0,0)} \frac{\sin x}{x}\right)\\ &= (1)(1)\\ &= 1.\end{aligned}$$

We have found that $\lim\limits_{(x,y)\to(0,0)} \frac{\cos y \sin x}{x} = f(0,0)$, so f is continuous at $(0,0)$.

A similar analysis shows that f is continuous at all points in $\mathbb{R}^2$. As long as $x \neq 0$, we can evaluate the limit directly; when $x = 0$, a similar analysis shows that the limit is $\cos y$. Thus we can say that f is continuous everywhere. A graph of f is given in Figure 12.10. Notice how it has no breaks, jumps, etc.

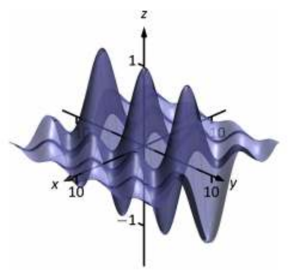


Figure 12.10: A graph of $f(x,y)$ in Example 12.2.6.

The following theorem is very similar to Theorem 8, giving us ways to combine continuous functions to create other continuous functions.

THEOREM 102 Properties of Continuous Functions

Let f and g be continuous on an open disk B, let c be a real number, and let n be a positive integer. The following functions are continuous on B.

1. Sums/Differences: $f \pm g$
2. Constant Multiples: $c \cdot f$
3. Products: $f \cdot g$
4. Quotients: f/g (as longs as $g \neq 0$ on B)
5. Powers: f^n
6. Roots: $\sqrt[n]{f}$ (if n is even then $f \geq 0$ on B; if n is odd, then true for all values of f on B.)
7. Compositions: Adjust the definitions of f and g to: Let f be continuous on B, where the range of f on B is J, and let g be a single variable function that is continuous on J. Then $g \circ f$, i.e., $g(f(x,y))$, is continuous on B.

Example 12.2.7: Establishing continuity of a function

Let $f(x,y) = \sin(x^2 \cos y)$. Show f is continuous everywhere.

Solution

We will apply both Theorems 8 and 102. Let $f_1(x,y) = x^2$. Since y is not actually used in the function, and polynomials are continuous (by Theorem 8), we conclude f_1 is continuous everywhere. A similar statement can be made about $f_2(x,y) = \cos y$.

> Part 3 of Theorem 102 states that $f_3 = f_1 \cdot f_2$ is continuous everywhere, and Part 7 of the theorem states the composition of sine with f_3 is continuous: that is, $\sin(f_3) = \sin(x^2 \cos y)$ is continuous everywhere.

Functions of Three Variables

The definitions and theorems given in this section can be extended in a natural way to definitions and theorems about functions of three (or more) variables. We cover the key concepts here; some terms from Definitions 79 and 81 are not redefined but their analogous meanings should be clear to the reader.

Definition 82 Open Balls, Limit, Continuous

1. An **open ball** in $\mathbb{R}^3$ centered at (x_0, y_0, z_0) with radius r is the set of all points (x, y, z) such that $\sqrt{(x-x_0)^2+(y-y_0)^2+(z-z_0)^2} = r$.
2. Let D be an open set in $\mathbb{R}^3$ containing (x_0, y_0, z_0), and let $f(x, y, z)$ be a function of three variables defined on D, except possibly at (x_0, y_0, z_0). The *limit* of $f(x, y, z)$ as (x, y, z) approaches (x_0, y_0, z_0) is L, denoted

$$\lim_{(x,y,z)\to(x_0,y_0,z_0)} f(x,y,z) = L, \tag{12.2.9}$$

means that given any $\epsilon > 0$, there is a $\delta > 0$ such that for all $(x, y, z) \neq (x_0, y_0, z_0)$, if (x, y, z) is in the open ball centered at (x_0, y_0, z_0) with radius δ, then $|f(x, y, z) - L| < \epsilon$.
3. Let $f(x, y, z)$ be defined on an open ball B containing (x_0, y_0, z_0). f is **continuous** at (x_0, y_0, z_0) if $\lim\limits_{(x,y,z)\to(x_0,y_0,z_0)} f(x,y,z) = f(x_0, y_0, z_0)$.

These definitions can also be extended naturally to apply to functions of four or more variables. Theorem 102 also applies to function of three or more variables, allowing us to say that the function

$$f(x,y,z) = \frac{e^{x^2+y}\sqrt{y^2+z^2+3}}{\sin(xyz)+5} \tag{12.2.10}$$

is continuous everywhere.

When considering single variable functions, we studied limits, then continuity, then the derivative. In our current study of multivariable functions, we have studied limits and continuity. In the next section we study derivation, which takes on a slight twist as we are in a multivarible context.

This page titled 12.2: Limits and Continuity of Multivariable Functions is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

12.3: Partial Derivatives

Let y be a function of x. We have studied in great detail the derivative of y with respect to x, that is, $\frac{dy}{dx}$, which measures the rate at which y changes with respect to x. Consider now $z = f(x, y)$. It makes sense to want to know how z changes with respect to x and/or y. This section begins our investigation into these rates of change.

Consider the function $z = f(x, y) = x^2 + 2y^2$, as graphed in Figure 12.11(a). By fixing $y = 2$, we focus our attention to all points on the surface where the y-value is 2, shown in both parts (a) and (b) of the figure. These points form a curve in space: $z = f(x, 2) = x^2 + 8$ which is a function of just one variable. We can take the derivative of z with respect to x along this curve and find equations of tangent lines, etc.

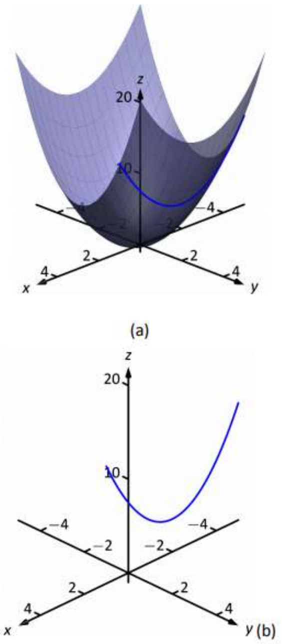


Figure 12.11: By fixing y=2, the surface $f(x, y) = x^2 + 2y^2$ is a curve in space.

The key notion to extract from this example is: by treating y as constant (it does not vary) we can consider how z changes with respect to x. In a similar fashion, we can hold x constant and consider how z changes with respect to y. This is the underlying principle of **partial derivatives**. We state the formal, limit--based definition first, then show how to compute these partial derivatives without directly taking limits.

Definition 83 Partial Derivative

Let $z = f(x, y)$ be a continuous function on an open set S in $\mathbb{R}^2$.

1. The **partial derivative of f with respect to x** is:

$$f_x(x, y) = \lim_{h \to 0} \frac{f(x + h, y) - f(x, y)}{h}. \tag{12.3.1}$$

2. The **partial derivative of f with respect to y** is:

$$f_y(x, y) = \lim_{h \to 0} \frac{f(x, y + h) - f(x, y)}{h}. \tag{12.3.2}$$

Note: Alternate notations for $f_x(x, y)$ include:

$$\frac{\partial}{\partial x} f(x, y), \ \frac{\partial f}{\partial x}, \ \frac{\partial z}{\partial x}, \ \text{and } z_x, \tag{12.3.3}$$

with similar notations for $f_y(x, y)$. For ease of notation, $f_x(x, y)$ is often abbreviated f_x.

Example 12.3.1: Computing partial derivatives with the limit definition

Let $f(x,y)=x^2y+2x+y^3$. Find $f_x(x,y)$ using the limit definition.

Solution

Using Definition 83, we have:

$$\begin{aligned} f_x(x,y) &= \lim_{h\to 0}\frac{f(x+h,y)-f(x,y)}{h} \\ &= \lim_{h\to 0}\frac{(x+h)^2y+2(x+h)+y^3-(x^2y+2x+y^3)}{h} \\ &= \lim_{h\to 0}\frac{(x^2y+2xhy+h^2y+2x+2h+y^3-(x^2y+2x+y^3)}{h} \\ &= \lim_{h\to 0}\frac{2xhy+h^2y+2h}{h} \\ &= \lim_{h\to 0} 2xy+hy+2 \\ &= 2xy+2. \end{aligned}$$

We have found $f_x(x,y)=2xy+2$.

Example 12.3.1 found a partial derivative using the formal, limit--based definition. Using limits is not necessary, though, as we can rely on our previous knowledge of derivatives to compute partial derivatives easily. When computing $f_x(x,y)$, we hold y fixed -- it does not vary. Therefore we can compute the derivative with respect to x by treating y as a constant or coefficient.

Just as $\frac{d}{dx}\left(5x^2\right)=10x$, we compute $\frac{\partial}{\partial x}\left(x^2y\right)=2xy$. Here we are treating y as a coefficient.

Just as $\frac{d}{dx}\left(5^3\right)=0$, we compute $\frac{\partial}{\partial x}\left(y^3\right)=0$. Here we are treating y as a constant. More examples will help make this clear.

Example 12.3.2: Finding partial derivatives

Find $f_x(x,y)$ and $f_y(x,y)$ in each of the following.

1. $f(x,y)=x^3y^2+5y^2-x+7$
2. $f(x,y)=\cos(xy^2)+\sin x$
3. $f(x,y)=e^{x^2y^3}\sqrt{x^2+1}$

Solution

1. We have $f(x,y)=x^3y^2+5y^2-x+7$. Begin with $f_x(x,y)$. Keep y fixed, treating it as a constant or coefficient, as appropriate:

   $$f_x(x,y)=3x^2y^2-1. \tag{12.3.4}$$

   Note how the $5y^2$ and 7 terms go to zero. To compute $f_y(x,y)$, we hold x fixed:

   $$f_y(x,y)=2x^3y+10y. \tag{12.3.5}$$

   Note how the $-x$ and 7 terms go to zero.
2. We have $f(x,y)=\cos(xy^2)+\sin x$.
   Begin with $f_x(x,y)$. We need to apply the Chain Rule with the cosine term; y^2 is the coefficient of the x-term inside the cosine function.

   $$f_x(x,y)=-\sin(xy^2)(y^2)+\cos x=-y^2\sin(xy^2)+\cos x. \tag{12.3.6}$$

   To find $f_y(x,y)$, note that x is the coefficient of the y^2 term inside of the cosine term; also note that since x is fixed, $\sin x$ is also fixed, and we treat it as a constant.

   $$f_y(x,y)=-\sin(xy^2)(2xy)=-2xy\sin(xy^2). \tag{12.3.7}$$

3. We have $f(x,y)=e^{x^2y^3}\sqrt{x^2+1}$.
   Beginning with $f_x(x,y)$, note how we need to apply the Product Rule.

$$\begin{aligned} f_x(x,y) &= e^{x^2y^3}(2xy^3)\sqrt{x^2+1}+e^{x^2y^3}\frac{1}{2}\left(x^2+1\right)^{-1/2}(2x) \\ &= 2xy^3e^{x^2y^3}\sqrt{x^2+1}+\frac{xe^{x^2y^3}}{\sqrt{x^2+1}}. \end{aligned}$$

Note that when finding $f_y(x,y)$ we do not have to apply the Product Rule; since $\sqrt{x^2+1}$ does not contain y, we treat it as fixed and hence becomes a coefficient of the $e^{x^2y^3}$ term.

$$f_y(x,y) = e^{x^2y^3}(3x^2y^2)\sqrt{x^2+1} = 3x^2y^2e^{x^2y^3}\sqrt{x^2+1}. \tag{12.3.8}$$

We have shown *how* to compute a partial derivative, but it may still not be clear what a partial derivative *means*. Given $z=f(x,y)$, $f_x(x,y)$ measures the rate at which z changes as only x varies: y is held constant.

Imagine standing in a rolling meadow, then beginning to walk due east. Depending on your location, you might walk up, sharply down, or perhaps not change elevation at all. This is similar to measuring z_x: you are moving only east (in the "x"-direction) and not north/south at all. Going back to your original location, imagine now walking due north (in the "y"-direction). Perhaps walking due north does not change your elevation at all. This is analogous to $z_y=0$: z does not change with respect to y. We can see that z_x and z_y do not have to be the same, or even similar, as it is easy to imagine circumstances where walking east means you walk downhill, though walking north makes you walk uphill.

The following example helps us visualize this more.

Example 12.3.3: Evaluating partial derivatives

Let $z=f(x,y)=-x^2-\frac{1}{2}y^2+xy+10$. Find $f_x(2,1)$ and $f_y(2,1)$ and interpret their meaning.

Solution

We begin by computing $f_x(x,y)=-2x+y$ and $f_y(x,y)=-y+x$. Thus

$$f_x(2,1)=-3 \quad \text{and} \quad f_y(2,1)=1. \tag{12.3.9}$$

It is also useful to note that $f(2,1)=7.5$. What does each of these numbers mean?

Consider $f_x(2,1)=-3$, along with Figure 12.12(a). If one "stands" on the surface at the point $(2,1,7.5)$ and moves parallel to the x-axis (i.e., only the x-value changes, not the y-value), then the instantaneous rate of change is -3. Increasing the x-value will decrease the z-value; decreasing the x-value will increase the z-value.

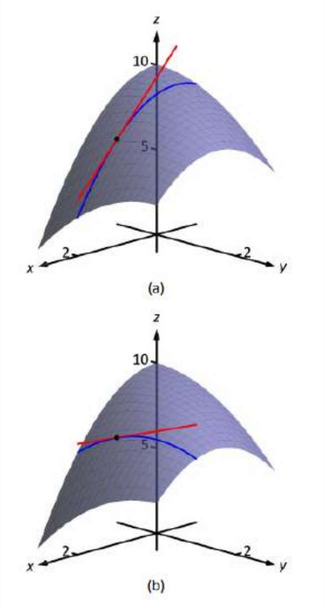


Figure 12.12: Illustrating the meaning of partial derivatives.

Now consider $f_y(2,1)=1$, illustrated in Figure 12.12(b). Moving along the curve drawn on the surface, i.e., parallel to the y-axis and not changing the x-values, increases the z-value instantaneously at a rate of 1. Increasing the y-value by 1 would increase the z-value by approximately 1.

Since the magnitude of f_x is greater than the magnitude of f_y at $(2,1)$, it is "steeper" in the x-direction than in the y-direction.

Second Partial Derivatives

Let $z=f(x,y)$. We have learned to find the partial derivatives $f_x(x,y)$ and $f_y(x,y)$, which are each functions of x and y. Therefore we can take partial derivatives of them, each with respect to x and y. We define these "second partials" along with the notation, give examples, then discuss their meaning.

Definition 84 Second Partial Derivative and Mixed Partial Derivative

Let $z=f(x,y)$ be continuous on an open set S.

1. The **second partial derivative of f with respect to x** then x is

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right)=\frac{\partial^2 f}{\partial x^2}=\left(f_x\right)_x=f_{xx} \tag{12.3.10}$$

2. The **second partial derivative of f with respect to x** then y is

$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)=\frac{\partial^2 f}{\partial y\partial x}=\left(f_x\right)_y=f_{xy} \tag{12.3.11}$$

Similar definitions hold for $\frac{\partial^2 f}{\partial y^2}=f_{yy}$ and $\frac{\partial^2 f}{\partial x\partial y}=f_{yx}$.

The second partial derivatives f_{xy} and f_{yx} are **mixed partial derivatives**.

The notation of second partial derivatives gives some insight into the notation of the second derivative of a function of a single variable. If $y=f(x)$, then $f''(x)=\frac{d^2y}{dx^2}$. The "d^2y" portion means "take the derivative of y twice," while "dx^2" means "with respect to x both times." When we only know of functions of a single variable, this latter phrase seems silly: there is only one variable to take the derivative with respect to. Now that we understand functions of multiple variables, we see the importance of specifying which variables we are referring to.

Note: The terms in Definition 84 all depend on limits, so each definition comes with the caveat "where the limit exists."

Example 12.3.4: Second partial derivatives

For each of the following, find all six first and second partial derivatives. That is, find

$$f_x,\quad f_y,\quad f_{xx},\quad f_{yy},\quad f_{xy}\quad \text{and}\quad f_{yx}\,. \tag{12.3.12}$$

1. $f(x,y)=x^3y^2+2xy^3+\cos x$
2. $f(x,y)=\frac{x^3}{y^2}$
3. $f(x,y)=e^x\sin(x^2y)$

Solution

In each, we give f_x and f_y immediately and then spend time deriving the second partial derivatives.

1. $f(x,y)=x^3y^2+2xy^3+\cos x$

$$\begin{aligned}
f_x(x,y)&=3x^2y^2+2y^3-\sin x\\
f_y(x,y)&=2x^3y+6xy^2\\
f_{xx}(x,y)&=\frac{\partial}{\partial x}\left(f_x\right)=\frac{\partial}{\partial x}\left(3x^2y^2+2y^3-\sin x\right)=6xy^2-\cos x\\
f_{yy}(x,y)&=\frac{\partial}{\partial y}\left(f_y\right)=\frac{\partial}{\partial y}\left(2x^3y+6xy^2\right)=2x^3+12xy\\
f_{xy}(x,y)&=\frac{\partial}{\partial y}\left(f_x\right)=\frac{\partial}{\partial y}\left(3x^2y^2+2y^3-\sin x\right)=6x^2y+6y^2\\
f_{yx}(x,y)&=\frac{\partial}{\partial x}\left(f_x\right)=\frac{\partial}{\partial x}\left(2x^3y+6xy^2\right)=6x^2y+6y^2
\end{aligned}$$

2. $f(x,y) = \frac{x^3}{y^2} = x^3y^{-2}$
   $f_x(x,y) = \frac{3x^2}{y^2}$
   $f_y(x,y) = -\frac{2x^3}{y^3}$
   $f_{xx}(x,y) = \frac{\partial}{\partial x}\left(f_x\right) = \frac{\partial}{\partial x}\left(\frac{3x^2}{y^2}\right) = \frac{6x}{y^2}$
   $f_{yy}(x,y) = \frac{\partial}{\partial y}\left(f_y\right) = \frac{\partial}{\partial y}\left(-\frac{2x^3}{y^3}\right) = \frac{6x^3}{y^4}$
   $f_{xy}(x,y) = \frac{\partial}{\partial y}\left(f_x\right) = \frac{\partial}{\partial y}\left(\frac{3x^2}{y^2}\right) = -\frac{6x^2}{y^3}$
   $f_{yx}(x,y) = \frac{\partial}{\partial x}\left(f_x\right) = \frac{\partial}{\partial x}\left(-\frac{2x^3}{y^3}\right) = -\frac{6x^2}{y^3}$
3. $f(x,y) = e^x\sin(x^2y)$
   Because the following partial derivatives get rather long, we omit the extra notation and just give the results. In several cases, multiple applications of the Product and Chain Rules will be necessary, followed by some basic combination of like terms.
   $f_x(x,y) = e^x\sin(x^2y) + 2xye^x\cos(x^2y)$
   $f_y(x,y) = x^2e^x\cos(x^2y)$
   $f_{xx}(x,y) = e^x\sin(x^2y) + 4xye^x\cos(x^2y) + 2ye^x\cos(x^2y) - 4x^2y^2e^x\sin(x^2y)$
   $f_{yy}(x,y) = -x^4e^x\sin(x^2y)$
   $f_{xy}(x,y) = x^2e^x\cos(x^2y) + 2xe^x\cos(x^2y) - 2x^3ye^x\sin(x^2y)$
   $f_{yx}(x,y) = x^2e^x\cos(x^2y) + 2xe^x\cos(x^2y) - 2x^3ye^x\sin(x^2y)$

Notice how in each of the three functions in Example 12.3.4, $f_{xy} = f_{yx}$. Due to the complexity of the examples, this likely is not a coincidence. The following theorem states that it is not.

theorem 103 Mixed Partial Derivatives

Let f be defined such that f_{xy} and f_{yx} are continuous on an open set S. Then for each point (x,y) in S, $f_{xy}(x,y) = f_{yx}(x,y)$.

Finding f_{xy} and f_{yx} independently and comparing the results provides a convenient way of checking our work.

Understanding Second Partial Derivatives

Now that we know *how* to find second partials, we investigate *what* they tell us.

Again we refer back to a function $y = f(x)$ of a single variable. The second derivative of f is "the derivative of the derivative," or "the rate of change of the rate of change." The second derivative measures how much the derivative is changing. If $f''(x) < 0$, then the derivative is getting smaller (so the graph of f is concave down); if $f''(x) > 0$, then the derivative is growing, making the graph of f concave up.

Now consider $z = f(x,y)$. Similar statements can be made about f_{xx} and f_{yy} as could be made about $f''(x)$ above. When taking derivatives with respect to x twice, we measure how much f_x changes with respect to x. If $f_{xx}(x,y) < 0$, it means that as x increases, f_x decreases, and the graph of f will be concave down *in the x-direction*. Using the analogy of standing in the rolling meadow used earlier in this section, f_{xx} measures whether one's path is concave up/down when walking due east.

Similarly, f_{yy} measures the concavity in the y-direction. If $f_{yy}(x,y) > 0$, then f_y is increasing with respect to y and the graph of f will be concave up in the y-direction. Appealing to the rolling meadow analogy again, f_{yy} measures whether one's path is concave up/down when walking due north.

We now consider the mixed partials f_{xy} and f_{yx}. The mixed partial f_{xy} measures how much f_x changes with respect to y. Once again using the rolling meadow analogy, f_x measures the slope if one walks due east. Looking east, begin walking *north* (side--stepping). Is the path towards the east getting steeper? If so, $f_{xy} > 0$. Is the path towards the east not changing in steepness? If so, then $f_{xy} = 0$. A similar thing can be said about f_{yx}: consider the steepness of paths heading north while side--stepping to the east.

The following example examines these ideas with concrete numbers and graphs.

Example 12.3.5: Understanding second partial derivatives

Let $z = x^2 - y^2 + xy$. Evaluate the 6 first and second partial derivatives at $(-1/2, 1/2)$ and interpret what each of these numbers mean.

Solution

We find that:

$f_x(x,y) = 2x + y$,\quad $f_y(x,y) = -2y + x$,\quad $f_{xx}(x,y) = 2$, \quad $f_{yy}(x,y) = -2$ and $f_{xy}(x,y) = f_{yx}(x,y) = 1$. Thus at $(-1/2, 1/2)$ we have

$$f_x(-1/2,1/2) = -1/2, \qquad f_y(-1/2,1/2) = -3/2. \tag{12.3.13}$$

The slope of the tangent line at $(-1/2, 1/2, -1/4)$ in the direction of x is $-1/2$: if one moves from that point parallel to the x-axis, the instantaneous rate of change will be $-1/2$. The slope of the tangent line at this point in the direction of y is $-3/2$: if one moves from this point parallel to the y-axis, the instantaneous rate of change will be $-3/2$. These tangents lines are graphed in Figure 12.13(a) and (b), respectively, where the tangent lines are drawn in a solid line.

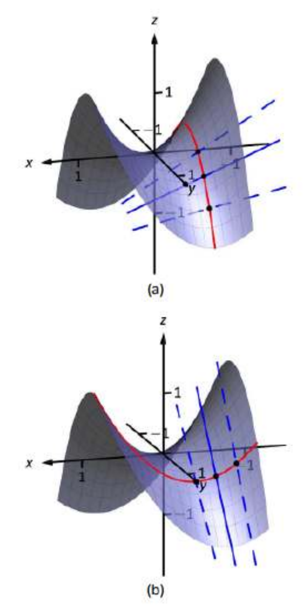


Figure 12.13: Understanding the second partial derivatives in Example 12.3.5.

Now consider only Figure 12.13(a). Three directed tangent lines are drawn (two are dashed), each in the direction of x; that is, each has a slope determined by f_x. Note how as y increases, the slope of these lines get closer to 0. Since the slopes are all negative, getting closer to 0 means the *slopes are increasing*. The slopes given by f_x are increasing as y increases, meaning f_{xy} must be positive.

Since $f_{xy} = f_{yx}$, we also expect f_y to increase as x increases. Consider Figure 12.13(b) where again three directed tangent lines are drawn, this time each in the direction of y with slopes determined by f_y. As x increases, the slopes become less steep (closer to 0). Since these are negative slopes, this means the slopes are increasing.

Thus far we have a visual understanding of f_x, f_y, and $f_{xy} = f_{yx}$. We now interpret f_{xx} and f_{yy}. In Figure 12.13(a), we see a curve drawn where x is held constant at $x = -1/2$: only y varies. This curve is clearly concave down, corresponding to the fact that $f_{yy} < 0$. In part (b) of the figure, we see a similar curve where y is constant and only x varies. This curve is concave up, corresponding to the fact that $f_{xx} > 0$.

Partial Derivatives and Functions of Three Variables

The concepts underlying partial derivatives can be easily extend to more than two variables. We give some definitions and examples in the case of three variables and trust the reader can extend these definitions to more variables if needed.

Definition 85 Partial Derivatives with Three Variables

Let $w = f(x,y,z)$ be a continuous function on an open set S in $\mathbb{R}^3$.

The **partial derivative of f with respect to** x is:

$$f_x(x,y,z) = \lim_{h\to 0} \frac{f(x+h,y,z) - f(x,y,z)}{h}. \tag{12.3.14}$$

Similar definitions hold for $f_y(x,y,z)$ and $f_z(x,y,z)$.

By taking partial derivatives of partial derivatives, we can find second partial derivatives of f with respect to z then y, for instance, just as before.

Example 12.3.6: Partial derivatives of functions of three variables

For each of the following, find f_x, f_y, f_z, f_{xz}, f_{yz}, and f_{zz}.

1. $f(x,y,z) = x^2y^3z^4 + x^2y^2 + x^3z^3 + y^4z^4$
2. $f(x,y,z) = x\sin(yz)$

Solution

1. $f_x = 2xy^3z^4 + 2xy^2 + 3x^2z^3; \quad f_y = 3x^2y^2z^4 + 2x^2y + 4y^3z^4;$
   $f_z = 4x^2y^3z^3 + 3x^3z^2 + 4y^4z^3; \quad f_{xz} = 8xy^3z^3 + 9x^2z^2;$
   $f_{yz} = 12x^2y^2z^3 + 16y^3z^3; \quad f_{zz} = 12x^2y^3z^2 + 6x^3z + 12y^4z^2$
2. $f_x = \sin(yz); \quad f_y = xz\cos(yz); \quad f_z = xy\cos(yz);$
   $f_{xz} = y\cos(yz); \quad f_{yz} = x\cos(yz) - xyz\sin(yz); \quad f_{zz} = -xy^2\sin(xy)$

Higher Order Partial Derivatives

We can continue taking partial derivatives of partial derivatives of partial derivatives of ...; we do not have to stop with second partial derivatives. These higher order partial derivatives do not have a tidy graphical interpretation; nevertheless they are not hard to compute and worthy of some practice. We do not formally define each higher order derivative, but rather give just a few examples of the notation.

$$f_{xyx}(x,y) = \frac{\partial}{\partial x}\left(\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)\right) \quad \text{and} \tag{12.3.15}$$

$$f_{xyz}(x,y,z) = \frac{\partial}{\partial z}\left(\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)\right). \tag{12.3.16}$$

Example 12.3.7: Higher order partial derivatives

a. Let $f(x,y) = x^2y^2 + \sin(xy)$. Find f_{xxy} and f_{yxx}.
b. Let $f(x,y,z) = x^3e^{xy} + \cos(z)$. Find f_{xyz}.

Solution

a. To find f_{xxy}, we first find f_x, then f_{xx}, then f_{xxy}:

$$\begin{aligned} f_x &= 2xy^2 + y\cos(xy) \qquad f_{xx} = 2y^2 - y^2\sin(xy) \\ f_{xxy} &= 4y - 2y\sin(xy) - xy^2\cos(xy). \end{aligned}$$

To find f_{yxx}, we first find f_y, then f_{yx}, then f_{yxx}:

$$\begin{aligned} f_y &= 2x^2y + x\cos(xy) \qquad f_{yx} = 4xy + \cos(xy) - xy\sin(xy) \\ f_{yxx} &= 4y - y\sin(xy) - \big(y\sin(xy) + xy^2\cos(xy)\big) \\ &= 4y - 2y\sin(xy) - xy^2\cos(xy). \end{aligned}$$

Note how $f_{xxy} = f_{yxx}$.

b. To find f_{xyz}, we find f_x, then f_{xy}, then f_{xyz}:

$$\begin{aligned} f_x &= 3x^2e^{xy} + x^3ye^{xy} \qquad f_{xy} = 3x^3e^{xy} + x^3e^{xy} + x^4ye^{xy} = 4x^3e^{xy} + x^4ye^{xy} \\ f_{xyz} &= 0. \end{aligned}$$

In the previous example we saw that $f_{xxy} = f_{yxx}$; this is not a coincidence. While we do not state this as a formal theorem, as long as each partial derivative is continuous, it does not matter the order in which the partial derivatives are taken. For instance, $f_{xxy} = f_{xyx} = f_{yxx}$.

This can be useful at times. Had we known this, the second part of Example $12.3.7$ would have been much simpler to compute. Instead of computing f_{xyz} in the x, y then z orders, we could have applied the z, then x then y order (as $f_{xyz} = f_{zxy}$). It is easy to see that $f_z = -\sin z$; then f_{zx} and f_{zxy} are clearly 0 as f_z does not contain an x or y.

A brief review of this section: partial derivatives measure the instantaneous rate of change of a multivariable function with respect to one variable. With $z = f(x,y)$, the partial derivatives f_x and f_y measure the instantaneous rate of change of z when moving parallel to the x- and y-axes, respectively. How do we measure the rate of change at a point when we do not move parallel to one of these axes? What if we move in the direction given by the vector $\langle 2,1\rangle$? Can we measure that rate of change? The answer is, of course, yes, we can. This is the topic of Section 12.6. First, we need to define what it means for a function of two variables to be *differentiable*.

This page titled 12.3: Partial Derivatives is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

12.4: Differentiability and the Total Differential

We studied **differentials** in Section 4.4, where Definition 18 states that if $y = f(x)$ and f is differentiable, then $dy = f'(x)dx$. One important use of this differential is in Integration by Substitution. Another important application is approximation. Let $\Delta x = dx$ represent a change in x. When dx is small, $dy \approx \Delta y$, the change in y resulting from the change in x. Fundamental in this understanding is this: as dx gets small, the difference between Δy and dy goes to 0. Another way of stating this: as dx goes to 0, the *error* in approximating Δy with dy goes to 0.

We extend this idea to functions of two variables. Let $z = f(x, y)$, and let $\Delta x = dx$ and $\Delta y = dy$ represent changes in x and y, respectively. Let $\Delta z = f(x + dx, y + dy) - f(x, y)$ be the change in z over the change in x and y. Recalling that f_x and f_y give the instantaneous rates of z-change in the x- and y-directions, respectively, we can approximate Δz with $dz = f_x dx + f_y dy$; in words, the total change in z is approximately the change caused by changing x plus the change caused by changing y. In a moment we give an indication of whether or not this approximation is any good. First we give a name to dz.

Definition 86: Total Differential

Let $z = f(x, y)$ be continuous on an open set S. Let dx and dy represent changes in x and y, respectively. Where the partial derivatives f_x and f_y exist, the **total differential of** z is

$$dz = f_x(x, y)dx + f_y(x, y)dy. \tag{12.4.1}$$

Example 12.4.1: Finding the total differential

Let $z = x^4 e^{3y}$. Find dz.

Solution

We compute the partial derivatives: $f_x = 4x^3e^{3y}$ and $f_y = 3x^4e^{3y}$. Following Definition 86, we have

$$dz = 4x^3e^{3y}dx + 3x^4e^{3y}dy. \tag{12.4.2}$$

We *can* approximate Δz with dz, but as with all approximations, there is error involved. A good approximation is one in which the error is small. At a given point (x_0, y_0), let E_x and E_y be functions of dx and dy such that $E_x dx + E_y dy$ describes this error. Then

$$\begin{aligned}\Delta z &= dz + E_x dx + E_y dy \\ &= f_x(x_0, y_0)dx + f_y(x_0, y_0)dy + E_x dx + E_y dy.\end{aligned}$$

If the approximation of Δz by dz is good, then as dx and dy get small, so does $E_x dx + E_y dy$. The approximation of Δz by dz is even better if, as dx and dy go to 0, so do E_x and E_y. This leads us to our definition of differentiability.

Definition 87: Multivariable Differentiability

Let $z = f(x, y)$ be defined on an open set S containing (x_0, y_0) where $f_x(x_0, y_0)$ and $f_y(x_0, y_0)$ exist. Let dz be the total differential of z at (x_0, y_0), let $\Delta z = f(x_0 + dx, y_0 + dy) - f(x_0, y_0)$, and let E_x and E_y be functions of dx and dy such that

$$\Delta z = dz + E_x dx + E_y dy. \tag{12.4.3}$$

1. f is **differentiable at** (x_0, y_0) if, given $\epsilon > 0$, there is a $\delta > 0$ such that if $||\langle dx, dy\rangle|| < \delta$, then $||\langle E_x, E_y\rangle|| < \epsilon$. That is, as dx and dy go to 0, so do E_x and E_y.
2. f is **differentiable on** S if f is differentiable at every point in S. If f is differentiable on $\mathbb{R}^2$, we say that f is **differentiable everywhere**.

Example 12.4.2: Showing a function is differentiable

Show $f(x, y) = xy + 3y^2$ is differentiable using Definition 87.

Solution

We begin by finding $f(x + dx, y + dy)$, Δz, f_x and f_y.

$$\begin{aligned}f(x + dx, y + dy) &= (x + dx)(y + dy) + 3(y + dy)^2 \\ &= xy + xdy + ydx + dxdy + 3y^2 + 6ydy + 3dy^2.\end{aligned}$$

$\Delta z = f(x+dx, y+dy) - f(x,y)$, so

$$\Delta z = xdy + ydx + dxdy + 6ydy + 3dy^2. \tag{12.4.4}$$

It is straightforward to compute $f_x = y$ and $f_y = x + 6y$. Consider once more Δz:

$$\begin{aligned} \Delta z &= xdy + ydx + dxdy + 6ydy + 3dy^2 \qquad \text{(now reorder)} \\ &= ydx + xdy + 6ydy + dxdy + 3dy^2 \\ &= \underbrace{(y)}_{f_x} dx + \underbrace{(x+6y)}_{f_y} dy + \underbrace{(dy)}_{E_x} dx + \underbrace{(3dy)}_{E_y} dy \\ &= f_x dx + f_y dy + E_x dx + E_y dy. \end{aligned}$$

With $E_x = dy$ and $E_y = 3dy$, it is clear that as dx and dy go to 0, E_x and E_y also go to 0. Since this did not depend on a specific point (x_0, y_0), we can say that $f(x,y)$ is differentiable for all pairs (x,y) in $\mathbb{R}^2$, or, equivalently, that f is differentiable everywhere.

Our intuitive understanding of differentiability of functions $y = f(x)$ of one variable was that the graph of f was "smooth." A similar intuitive understanding of functions $z = f(x,y)$ of two variables is that the surface defined by f is also "smooth," not containing cusps, edges, breaks, etc. The following theorem states that differentiable functions are continuous, followed by another theorem that provides a more tangible way of determining whether a great number of functions are differentiable or not.

THEOREM 104: Continuity and Differentiability of Multivariable Functions

Let $z = f(x,y)$ be defined on an open set S containing (x_0, y_0).

If f is differentiable at (x_0, y_0), then f is continuous at (x_0, y_0).

THEOREM 105: Differentiability of Multivariable Functions

Let $z = f(x,y)$ be defined on an open set S containing (x_0, y_0).

If f_x and f_y are both continuous on S, then f is differentiable on S.

The theorems assure us that essentially all functions that we see in the course of our studies here are differentiable (and hence continuous) on their natural domains. There is a difference between Definition 87 and Theorem 105, though: it is possible for a function f to be differentiable yet f_x and/or f_y is *not* continuous. Such strange behavior of functions is a source of delight for many mathematicians.

When f_x and f_y exist at a point but are not continuous at that point, we need to use other methods to determine whether or not f is differentiable at that point.

For instance, consider the function

$$f(x,y) = \begin{cases} \frac{xy}{x^2+y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases} \tag{12.4.5}$$

We can find $f_x(0,0)$ and $f_y(0,0)$ using Definition 83:

$$\begin{aligned} f_x(0,0) &= \lim_{h\to 0} \frac{f(0+h,0) - f(0,0)}{h} \\ &= \lim_{h\to 0} \frac{0}{h^2} = 0; \\ f_y(0,0) &= \lim_{h\to 0} \frac{f(0,0+h) - f(0,0)}{h} \\ &= \lim_{h\to 0} \frac{0}{h^2} = 0. \end{aligned}$$

Both f_x and f_y *exist* at $(0,0)$, but they are not continuous at $(0,0)$, as

$$f_x(x,y) = \frac{y(y^2-x^2)}{(x^2+y^2)^2} \qquad \text{and} \qquad f_y(x,y) = \frac{x(x^2-y^2)}{(x^2+y^2)^2} \tag{12.4.6}$$

are not continuous at $(0,0)$. (Take the limit of f_x as $(x,y) \to (0,0)$ along the x- and y-axes; they give different results.) So even though f_x and f_y *exist* at every point in the x-y plane, they are not continuous. Therefore it is possible, by Theorem 105, for f to not be differentiable.

Indeed, it is not. One can show that f is not continuous at $(0,0)$ (see Example 12.2.4), and by Theorem 104, this means f is not differentiable at $(0,0)$.

Approximating with the Total Differential

By the definition, when f is differentiable dz is a good approximation for Δz when dx and dy are small. We give some simple examples of how this is used here.

Example 12.4.3: Approximating with the total differential

Let $z=\sqrt{x}\sin y$. Approximate $f(4.1,0.8)$.

Solution

Recognizing that $\pi/4\ \approx 0.785\ \approx 0.8$, we can approximate $f(4.1,0.8)$ using $f(4,\pi/4)$. We can easily compute $f(4,\pi/4)$ $=\sqrt{4}\sin(\pi/4)=2\left(\frac{\sqrt{2}}{2}\right)=\sqrt{2}\approx 1.414$. Without calculus, this is the best approximation we could reasonably come up with. The total differential gives us a way of adjusting this initial approximation to hopefully get a more accurate answer.

We let $\Delta z=f(4.1,0.8)-f(4,\pi/4)$. The total differential dz is approximately equal to Δz, so

$$f(4.1,0.8)-f(4,\pi/4)\approx dz \quad \Rightarrow \quad f(4.1,0.8)\approx dz+f(4,\pi/4). \tag{12.4.7}$$

To find dz, we need f_x and f_y.

$$\begin{aligned} f_x(x,y)=\frac{\sin y}{2\sqrt{x}} \quad \Rightarrow \quad f_x(4,\pi/4)&=\frac{\sin\pi/4}{2\sqrt{4}}\\ &=\frac{\sqrt{2}/2}{4}=\sqrt{2}/8.\\ f_y(x,y)=\sqrt{x}\cos y \quad \Rightarrow \quad f_y(4,\pi/4)&=\sqrt{4}\frac{\sqrt{2}}{2}\\ &=\sqrt{2}. \end{aligned}$$

Approximating 4.1 with 4 gives $dx\ =0.1$; approximating 0.8 with $\pi/4$ gives $dy\approx 0.015$. Thus

$$\begin{aligned} dz(4,\pi/4)&=f_x(4,\pi/4)(0.1)+f_y(4,\pi/4)(0.015)\\ &=\frac{\sqrt{2}}{8}(0.1)+\sqrt{2}(0.015)\\ &\approx 0.039. \end{aligned}$$

Returning to Equation 12.4.7, we have

$$f(4.1,0.8)\approx 0.039+1.414=1.4531. \tag{12.4.8}$$

We, of course, can compute the actual value of $f(4.1,0.8)$ with a calculator; the actual value, accurate to 5 places after the decimal, is 1.45254. Obviously our approximation is quite good.

The point of the previous example was *not* to develop an approximation method for known functions. After all, we can very easily compute $f(4.1,0.8)$ using readily available technology. Rather, it serves to illustrate how well this method of approximation works, and to reinforce the following concept:

$$\text{"New position} = \text{old position} + \text{amount of change,” so} \tag{12.4.9}$$

$$\text{"New position} \approx \text{old position} + \text{approximate amount of change.”} \tag{12.4.10}$$

In the previous example, we could easily compute $f(4,\pi/4)$ and could approximate the amount of z-change when computing $f(4.1,0.8)$, letting us approximate the new z-value.

It may be surprising to learn that it is not uncommon to know the values of f, f_x and f_y at a particular point without actually knowing the function f. The total differential gives a good method of approximating f at nearby points.

Example 12.4.4: Approximating an unknown function

Given that $f(2,-3)=6$, $f_x(2,-3)=1.3$ and $f_y(2,-3)=-0.6$, approximate $f(2.1,-3.03)$.

Solution

The total differential approximates how much f changes from the point $(2,-3)$ to the point $(2.1,-3.03)$. With $dx=0.1$ and $dy=-0.03$, we have

$$\begin{aligned} dz &= f_x(2,-3)dx + f_y(2,-3)dy \\ &= 1.3(0.1)+(-0.6)(-0.03) \\ &= 0.148. \end{aligned}$$

The change in z is approximately 0.148, so we approximate $f(2.1,-3.03)\approx 6.148.$

Error/Sensitivity Analysis

The total differential gives an approximation of the change in z given small changes in x and y. We can use this to approximate error propagation; that is, if the input is a little off from what it should be, how far from correct will the output be? We demonstrate this in an example.

Example 12.4.5: Sensitivity analysis

A cylindrical steel storage tank is to be built that is 10ft tall and 4ft across in diameter. It is known that the steel will expand/contract with temperature changes; is the overall volume of the tank more sensitive to changes in the diameter or in the height of the tank?

Solution

A cylindrical solid with height h and radius r has volume $V=\pi r^2h$. We can view V as a function of two variables, r and h. We can compute partial derivatives of V:

$$\frac{\partial V}{\partial r}=V_r(r,h)=2\pi rh \qquad \text{and} \qquad \frac{\partial V}{\partial h}=V_h(r,h)=\pi r^2. \tag{12.4.11}$$

The total differential is $dV=(2\pi rh)dr+(\pi r^2)dh.$ When $h=10$ and $r=2$, we have $dV=40\pi dr+4\pi dh.$

Note that the coefficient of dr is $40\pi\approx 125.7$; the coefficient of dh is a tenth of that, approximately 12.57. A small change in radius will be multiplied by 125.7, whereas a small change in height will be multiplied by 12.57. Thus the volume of the tank is more sensitive to changes in radius than in height.

The previous example showed that the volume of a particular tank was more sensitive to changes in radius than in height. Keep in mind that this analysis only applies to a tank of those dimensions. A tank with a height of 1 ft and radius of 5 ft would be more sensitive to changes in height than in radius. One could make a chart of small changes in radius and height and find exact changes in volume given specific changes. While this provides exact numbers, it does not give as much insight as the error analysis using the total differential.

Differentiability of Functions of Three Variables

The definition of differentiability for functions of three variables is very similar to that of functions of two variables. We again start with the total differential.

Definition 88: Total Differential

Let $w=f(x,y,z)$ be continuous on an open set S. Let dx, dy and dz represent changes in x, y and z, respectively. Where the partial derivatives f_x, f_y and f_z exist, the **total differential of** w is

$$dz=f_x(x,y,z)dx+f_y(x,y,z)dy+f_z(x,y,z)dz. \tag{12.4.12}$$

This differential can be a good approximation of the change in w when $w=f(x,y,z)$ is **differentiable**.

Definition 89: Multivariable Differentiability

Let $w = f(x, y, z)$ be defined on an open ball B containing (x_0, y_0, z_0) where $f_x(x_0, y_0, z_0)$, $f_y(x_0, y_0, z_0)$ and $f_z(x_0, y_0, z_0)$ exist. Let dw be the total differential of w at (x_0, y_0, z_0), let $\Delta w = f(x_0 + dx, y_0 + dy, z_0 + dz) - f(x_0, y_0, z_0)$, and let E_x, E_y and E_z be functions of dx, dy and dz such that

$$\Delta w = dw + E_x dx + E_y dy + E_z dz \tag{12.4.13}$$

1. f is **differentiable at** (x_0, y_0, z_0) if, given $\epsilon > 0$, there is a $\delta > 0$ such that if $||\langle dx, dy, dz\rangle|| < \delta$, then $||\langle E_x, E_y, E_z\rangle|| < \epsilon$.
2. f is **differentiable on** B if f is differentiable at every point in B. If f is differentiable on $\mathbb{R}^3$, we say that f is **differentiable everywhere.**

Just as before, this definition gives a rigorous statement about what it means to be differentiable that is not very intuitive. We follow it with a theorem similar to Theorem 105.

THEOREM 106: Continuity and Differentiability of Functions of Three Variables

Let $w = f(x, y, z)$ be defined on an open ball B containing (x_0, y_0, z_0).

1. If f is differentiable at (x_0, y_0, z_0), then f is continuous at (x_0, y_0, z_0).
2. If f_x, f_y and f_z are continuous on B, then f is differentiable on B.

This set of definition and theorem extends to functions of any number of variables. The theorem again gives us a simple way of verifying that most functions that we encounter are differentiable on their natural domains.

This section has given us a formal definition of what it means for a functions to be "differentiable," along with a theorem that gives a more accessible understanding. The following sections return to notions prompted by our study of partial derivatives that make use of the fact that most functions we encounter are differentiable.

This page titled 12.4: Differentiability and the Total Differential is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

12.5: The Multivariable Chain Rule

The Chain Rule, as learned in Section 2.5, states that $\frac{d}{dx}\left(f\left(g(x)\right)\right)=f'\left(g(x)\right)g'(x)$. If $t=g(x)$, we can express the Chain Rule as

$$\frac{df}{dx}=\frac{df}{dt}\frac{dt}{dx}. \tag{12.5.1}$$

In this section we extend the Chain Rule to functions of more than one variable.

theorem 107 Multivariable Chain Rule, Part I

Let $z=f(x,y)$, $x=g(t)$ and $y=h(t)$, where f, g and h are differentiable functions. Then $z=f(x,y)=f\left(g(t),h(t)\right)$ is a function of t, and

$$\begin{aligned}\frac{dz}{dt}=\frac{df}{dt}&=f_x(x,y)\frac{dx}{dt}+f_y(x,y)\frac{dy}{dt}\\&=\frac{\partial f}{\partial x}\frac{dx}{dt}+\frac{\partial f}{\partial y}\frac{dy}{dt}.\end{aligned}$$

It is good to understand what the situation of $z=f(x,y)$, $x=g(t)$ and $y=h(t)$ describes. We know that $z=f(x,y)$ describes a surface; we also recognize that $x=g(t)$ and $y=h(t)$ are parametric equations for a curve in the x-y plane. Combining these together, we are describing a curve that lies on the surface described by f. The parametric equations for this curve are $x=g(t)$, $y=h(t)$ and $z=f\left(g(t),h(t)\right)$.

Consider Figure 12.14 in which a surface is drawn, along with a dashed curve in the x-y plane. Restricting f to just the points on this circle gives the curve shown on the surface. The derivative $\frac{df}{dt}$ gives the instantaneous rate of change of f with respect to t. If we consider an object traveling along this path, $\frac{df}{dt}$ gives the rate at which the object rises/falls.

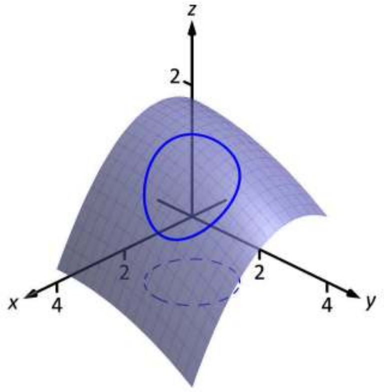


Figure 12.14: Understanding the application of the Multivariable Chain Rule.

We now practice applying the Multivariable Chain Rule.

Example $12.5.1$: Using the Multivariable Chain Rule

Let $z=x^2y+x$, where $x=\sin t$ and $y=e^{5t}$. Find $\frac{dz}{dt}$ using the Chain Rule.

Solution

Following Theorem 107, we find

$$f_x(x,y)=2xy+1,\qquad f_y(x,y)=x^2,\qquad \frac{dx}{dt}=\cos t,\qquad \frac{dy}{dt}=5e^{5t}. \tag{12.5.2}$$

Applying the theorem, we have

$$\frac{dz}{dt}=(2xy+1)\cos t+5x^2e^{5t}. \tag{12.5.3}$$

This may look odd, as it seems that $\frac{dz}{dt}$ is a function of x, y and t. Since x and y are functions of t, $\frac{dz}{dt}$ is really just a function of t, and we can replace x with $\sin t$ and y with e^{5t}:

$$\frac{dz}{dt} = (2xy+1)\cos t + 5x^2e^{5t} = (2\sin(t)e^{5t}+1)\cos t + 5e^{5t}\sin^2 t. \tag{12.5.4}$$

The previous example can make us wonder: if we substituted for x and y at the end to show that $\frac{dz}{dt}$ is really just a function of t, why not substitute before differentiating, showing clearly that z is a function of t?

That is, $z = x^2y + x = (\sin t)^2e^{5t} + \sin t$. Applying the Chain and Product Rules, we have

$$\frac{dz}{dt} = 2\sin t\cos t\, e^{5t} + 5\sin^2 t\, e^{5t} + \cos t, \tag{12.5.5}$$

which matches the result from the example.

This may now make one wonder "What's the point? If we could already find the derivative, why learn another way of finding it?" In some cases, applying this rule makes deriving simpler, but this is hardly the power of the Chain Rule. Rather, in the case where $z = f(x,y)$, $x = g(t)$ and $y = h(t)$, the Chain Rule is extremely powerful when *we do not know what* f, g and/or h are. It may be hard to believe, but often in "the real world" we know rate--of--change information (i.e., information about derivatives) without explicitly knowing the underlying functions. The Chain Rule allows us to combine several rates of change to find another rate of change. The Chain Rule also has theoretic use, giving us insight into the behavior of certain constructions (as we'll see in the next section).

We demonstrate this in the next example.

Example $12.5.2$: Applying the Multivarible Chain Rule

An object travels along a path on a surface. The exact path and surface are not known, but at time $t = t_0$ it is known that :

$$\frac{\partial z}{\partial x} = 5, \qquad \frac{\partial z}{\partial y} = -2, \qquad \frac{dx}{dt} = 3 \qquad \text{and} \qquad \frac{dy}{dt} = 7. \tag{12.5.6}$$

Find $\frac{dz}{dt}$ at time t_0.

Solution
The Multivariable Chain Rule states that

$$\begin{aligned}\frac{dz}{dt} &= \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}\\ &= 5(3) + (-2)(7)\\ &= 1.\end{aligned}$$

By knowing certain rates--of--change information about the surface and about the path of the particle in the x-y plane, we can determine how quickly the object is rising/falling.

We next apply the Chain Rule to solve a max/min problem.

Example $12.5.3$: Applying the Multivariable Chain Rule

Consider the surface $z = x^2 + y^2 - xy$, a paraboloid, on which a particle moves with x and y coordinates given by $x = \cos t$ and $y = \sin t$. Find $\frac{dz}{dt}$ when $t = 0$, and find where the particle reaches its maximum/minimum z-values.

Solution
It is straightforward to compute

$$f_x(x,y) = 2x - y, \qquad f_y(x,y) = 2y - x, \qquad \frac{dx}{dt} = -\sin t, \qquad \frac{dy}{dt} = \cos t. \tag{12.5.7}$$

Combining these according to the Chain Rule gives:

$$\frac{dz}{dt} = -(2x - y)\sin t + (2y - x)\cos t. \tag{12.5.8}$$

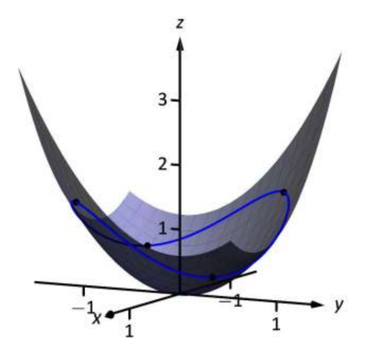


Figure 12.15: Plotting the path of a particle on a surface in Example 12.3.5

When $t = 0$, $x = 1$ and $y = 0$. Thus $\frac{dz}{dt} = -(2)(0) + (-1)(1) = -1$. When $t = 0$, the particle is moving down, as shown in Figure 12.15.

To find where z-value is maximized/minimized on the particle's path, we set $\frac{dz}{dt} = 0$ and solve for t:

$$\begin{aligned}\frac{dz}{dt} = 0 &= -(2x - y)\sin t + (2y - x)\cos t \\ 0 &= -(2\cos t - \sin t)\sin t + (2\sin t - \cos t)\cos t \\ 0 &= \sin^2 t - \cos^2 t \\ \cos^2 t &= \sin^2 t \\ t &= n\frac{\pi}{4} \quad (\text{for odd } n)\end{aligned}$$

We can use the First Derivative Test to find that on $[0, 2\pi]$, z has reaches its absolute minimum at $t = \pi/4$ and $5\pi/4$; it reaches its absolute maximum at $t = 3\pi/4$ and $7\pi/4$, as shown in Figure 12.15.

We can extend the Chain Rule to include the situation where z is a function of more than one variable, and each of these variables is also a function of more than one variable. The basic case of this is where $z = f(x, y)$, and x and y are functions of two variables, say s and t.

THEOREM 108 Multivariable Chain Rule, Part II

1. Let $z = f(x, y)$, $x = g(s, t)$ and $y = h(s, t)$, where f, g and h are differentiable functions. Then z is a function of s and t, and
   - $\frac{\partial z}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s}$, and
   - $\frac{\partial z}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t}$.
2. Let $z = f(x_1, x_2, \ldots, x_m)$ be a differentiable function of m variables, where each of the x_i is a differentiable function of the variables $t_1, t_2, \ldots, t_n$. Then z is a function of the t_i, and

$$\frac{\partial z}{\partial t_i} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial t_i} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial t_i} + \cdots + \frac{\partial f}{\partial x_m}\frac{\partial x_m}{\partial t_i}. \tag{12.5.9}$$

Example 12.5.4: Using the Multivarible Chain Rule, Part II

Let $z = x^2 y + x$, $x = s^2 + 3t$ and $y = 2s - t$. Find $\frac{\partial z}{\partial s}$ and $\frac{\partial z}{\partial t}$, and evaluate each when $s = 1$ and $t = 2$.

Solution

Following Theorem 108, we compute the following partial derivatives:

$$\frac{\partial f}{\partial x} = 2xy + 1 \qquad \frac{\partial f}{\partial y} = x^2, \tag{12.5.10}$$

$$\frac{\partial x}{\partial s}=2s \qquad \frac{\partial x}{\partial t}=3 \qquad \frac{\partial y}{\partial s}=2 \qquad \frac{\partial y}{\partial t}=-1. \tag{12.5.11}$$

Thus

$$\frac{\partial z}{\partial s}=(2xy+1)(2s)+(x^2)(2)=4xys+2s+2x^2, \quad \text{and} \tag{12.5.12}$$

$$\frac{\partial z}{\partial t}=(2xy+1)(3)+(x^2)(-1)=6xy-x^2+3. \tag{12.5.13}$$

When $s=1$ and $t=2$, $x=7$ and $y=0$, so

$$\frac{\partial z}{\partial s}=100 \qquad \text{and} \qquad \frac{\partial z}{\partial t}=-46. \tag{12.5.14}$$

Example 12.5.5: Using the Multivarible Chain Rule, Part II

Let $w=xy+z^2$, where $x=t^2e^s$, $y=t\cos s$, and $z=s\sin t$. Find $\frac{\partial w}{\partial t}$ when $s=0$ and $t=\pi$.

Solution
Following Theorem 108, we compute the following partial derivatives:

$$\frac{\partial f}{\partial x}=y \qquad \frac{\partial f}{\partial y}=x \qquad \frac{\partial f}{\partial z}=2z, \tag{12.5.15}$$

$$\frac{\partial x}{\partial t}=2te^s \qquad \frac{\partial y}{\partial t}=\cos s \qquad \frac{\partial z}{\partial t}=s\cos t. \tag{12.5.16}$$

Thus

$$\frac{\partial w}{\partial t}=y(2te^s)+x(\cos s)+2z(s\cos t). \tag{12.5.17}$$

When $s=0$ and $t=\pi$, we have $x=\pi^2$, $y=\pi$ and $z=0$. Thus

$$\frac{\partial w}{\partial t}=\pi(2\pi)+\pi^2=3\pi^2. \tag{12.5.18}$$

Implicit Differentiation

We studied finding $\frac{dy}{dx}$ when y is given as an implicit function of x in detail in Section 2.6. We find here that the Multivariable Chain Rule gives a simpler method of finding $\frac{dy}{dx}$.

For instance, consider the implicit function $x^2y-xy^3=3$. We learned to use the following steps to find $\frac{dy}{dx}$:

$$\begin{aligned}\frac{d}{dx}\left(x^2y-xy^3\right)&=\frac{d}{dx}\left(3\right)\\ 2xy+x^2\frac{dy}{dx}-y^3-3xy^2\frac{dy}{dx}&=0\\ \frac{dy}{dx}&=-\frac{2xy-y^3}{x^2-3xy^2}.\end{aligned} \tag{12.5.19}$$

Instead of using this method, consider $z=x^2y-xy^3$. The implicit function above describes the level curve $z=3$. Considering x and y as functions of x, the Multivariable Chain Rule states that

$$\frac{dz}{dx} = \frac{\partial z}{\partial x}\frac{dx}{dx} + \frac{\partial z}{\partial y}\frac{dy}{dx}. \tag{12.5.20}$$

Since z is constant (in our example, $z = 3$), $\frac{dz}{dx} = 0$. We also know $\frac{dx}{dx} = 1$. Equation 12.5.20 becomes

$$\begin{aligned} 0 &= \frac{\partial z}{\partial x}(1) + \frac{\partial z}{\partial y}\frac{dy}{dx} \quad \Rightarrow \\ \frac{dy}{dx} &= -\frac{\partial z}{\partial x}\bigg/\frac{\partial z}{\partial y} \\ &= -\frac{f_x}{f_y}. \end{aligned}$$

Note how our solution for $\frac{dy}{dx}$ in Equation 12.5.19 is just the partial derivative of z with respect to x, divided by the partial derivative of z with respect to y.

We state the above as a theorem.

THEOREM 109 Implicit Differentiation

Let f be a differentiable function of x and y, where $f(x, y) = c$ defines y as an implicit function of x, for some constant c. Then

$$\frac{dy}{dx} = -\frac{f_x(x,y)}{f_y(x,y)}. \tag{12.5.21}$$

We practice using Theorem 109 by applying it to a problem from Section 2.6.

Example 12.5.6: Implicit Differentiation

Given the implicitly defined function $\sin(x^2y^2) + y^3 = x + y$, find y'. Note: this is the same problem as given in Example 2.6.4 from Section 2.6, where the solution took about a full page to find.

Solution

Let $f(x, y) = \sin(x^2y^2) + y^3 - x - y$; the implicitly defined function above is equivalent to $f(x, y) = 0$. We find $\frac{dy}{dx}$ by applying Theorem 109. We find

$$f_x(x,y) = 2xy^2\cos(x^2y^2) - 1 \qquad \text{and} \qquad f_y(x,y) = 2x^2y\cos(x^2y^2) + 3y^2 - 1, \tag{12.5.22}$$

so

$$\frac{dy}{dx} = -\frac{2xy^2\cos(x^2y^2) - 1}{2x^2y\cos(x^2y^2) + 3y^2 - 1}, \tag{12.5.23}$$

which matches our solution from Example 2.6.4.

This page titled 12.5: The Multivariable Chain Rule is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

12.6: Directional Derivatives

Partial derivatives give us an understanding of how a surface changes when we move in the x and y directions. We made the comparison to standing in a rolling meadow and heading due east: the amount of rise/fall in doing so is comparable to f_x. Likewise, the rise/fall in moving due north is comparable to f_y. The steeper the slope, the greater in magnitude f_y. But what if we didn't move due north or east? What if we needed to move northeast and wanted to measure the amount of rise/fall? Partial derivatives alone cannot measure this. This section investigates **directional derivatives**, which do measure this rate of change. We begin with a definition.

Definition 90 Directional Derivatives

Let $z = f(x, y)$ be continuous on an open set S and let $\vec{u} = \langle u_1, u_2\rangle$ be a unit vector. For all points (x, y), the **directional derivative of f at (x, y) in the direction of $\vec{u}$** is

$$D_{\vec{u}}\, f(x, y) = \lim_{h\to 0} \frac{f(x + hu_1, y + hu_2) - f(x, y)}{h}. \tag{12.6.1}$$

The partial derivatives f_x and f_y are defined with similar limits, but only x or y varies with h, not both. Here both x and y vary with a weighted h, determined by a particular unit vector $\vec{u}$. This may look a bit intimidating but in reality it is not too difficult to deal with; it often just requires extra algebra. However, the following theorem reduces this algebraic load.

Theorem 110 Directional Derivatives

Let $z = f(x, y)$ be differentiable on an open set S containing (x_0, y_0), and let $\vec{u} = \langle u_1, u_2\rangle$ be a unit vector. The directional derivative of f at (x_0, y_0) in the direction of $\vec{u}$ is

$$D_{\vec{u}}\, f(x_0, y_0) = f_x(x_0, y_0)u_1 + f_y(x_0, y_0)u_2. \tag{12.6.2}$$

Example 12.6.1: Computing directional derivatives

Let $z = 14 - x^2 - y^2$ and let $P = (1, 2)$. Find the directional derivative of f, at P, in the following directions:

1. toward the point $Q = (3, 4)$,
2. in the direction of $\langle 2, -1\rangle$, and
3. toward the origin.

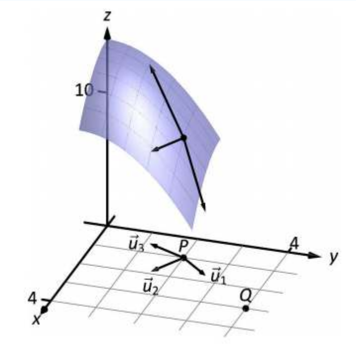


Figure 12.16: Understanding the directional derivative in Example 12.6.1

Solution

The surface is plotted in Figure 12.16, where the point $P = (1, 2)$ is indicated in the x, y-plane as well as the point $(1, 2, 9)$ which lies on the surface of f. We find that $f_x(x, y) = -2x$ and $f_x(1, 2) = -2$; $f_y(x, y) = -2y$ and $f_y(1, 2) = -4$.

1. Let $\vec{u}_1$ be the unit vector that points from the point $(1, 2)$ to the point $Q = (3, 4)$, as shown in the figure. The vector $\vec{PQ} = \langle 2, 2\rangle$; the unit vector in this direction is $\vec{u}_1 = \langle 1/\sqrt{2}, 1/\sqrt{2}\rangle$. Thus the directional derivative of f at $(1, 2)$ in the direction of $\vec{u}_1$ is

$$D_{\vec{u}_1} f(1, 2) = -2(1/\sqrt{2}) + (-4)(1/\sqrt{2}) = -6/\sqrt{2} \approx -4.24. \tag{12.6.3}$$

Thus the instantaneous rate of change in moving from the point $(1,2,9)$ on the surface in the direction of $\vec{u}_1$ (which points toward the point Q) is about -4.24. Moving in this direction moves one steeply downward.

2. We seek the directional derivative in the direction of $\langle 2,-1\rangle$. The unit vector in this direction is $\vec{u}_2=\langle 2/\sqrt{5},-1/\sqrt{5}\rangle$. Thus the directional derivative of f at $(1,2)$ in the direction of $\vec{u}_2$ is

$$D_{\vec{u}_2}f(1,2)=-2(2/\sqrt{5})+(-4)(-1/\sqrt{5})=0. \tag{12.6.4}$$

Starting on the surface of f at $(1,2)$ and moving in the direction of $\langle 2,-1\rangle$ (or $\vec{u}_2$) results in no instantaneous change in z-value. This is analogous to standing on the side of a hill and choosing a direction to walk that does not change the elevation. One neither walks up nor down, rather just "along the side" of the hill.

Finding these directions of "no elevation change" is important.

3. At $P=(1,2)$, the direction towards the origin is given by the vector $\langle -1,-2\rangle$; the unit vector in this direction is $\vec{u}_3$ $=\langle -1/\sqrt{5},-2/\sqrt{5}\rangle$. The directional derivative of f at P in the direction of the origin is

$$D_{\vec{u}_3}f(1,2)=-2(-1/\sqrt{5})+(-4)(-2/\sqrt{5})=10/\sqrt{5}\approx 4.47. \tag{12.6.5}$$

Moving towards the origin means "walking uphill" quite steeply, with an initial slope of about 4.47.

As we study directional derivatives, it will help to make an important connection between the unit vector $\vec{u}=\langle u_1,u_2\rangle$ that describes the direction and the partial derivatives f_x and f_y. We start with a definition and follow this with a Key Idea.

Definition 91 Gradient

Let $z=f(x,y)$ be differentiable on an open set S that contains the point (x_0,y_0).

1. The **gradient of** f is $\nabla f(x,y)=\langle f_x(x,y),f_y(x,y)\rangle$.
2. The **gradient of** f **at** (x_0,y_0) **is** $\nabla f(x_0,y_0)=\langle f_x(x_0,y_0),f_y(x_0,y_0)\rangle$.

Note: The symbol "∇" is named "nabla," derived from the Greek name of a Jewish harp. Oddly enough, in mathematics the expression ∇f is pronounced "del f."

To simplify notation, we often express the gradient as $\nabla f=\langle f_x,f_y\rangle$. The gradient allows us to compute directional derivatives in terms of a dot product.

KEY IDEA 55 The Gradient and Directional Derivatives

The directional derivative of $z=f(x,y)$ in the direction of $\vec{u}$ is

$$D_{\vec{u}}f=\nabla f\cdot\vec{u}. \tag{12.6.6}$$

The properties of the dot product previously studied allow us to investigate the properties of the directional derivative. Given that the directional derivative gives the instantaneous rate of change of z when moving in the direction of $\vec{u}$, three questions naturally arise:

1. In what direction(s) is the change in z the greatest (i.e., the "steepest uphill")?
2. In what direction(s) is the change in z the least (i.e., the "steepest downhill")?
3. In what direction(s) is there no change in z?

Using the key property of the dot product, we have

$$\nabla f\cdot\vec{u}=\|\nabla f\|\,\|u\|\cos\theta=\|\nabla f\|\cos\theta, \tag{12.6.7}$$

where θ is the angle between the gradient and $\vec{u}$. (Since $\vec{u}$ is a unit vector, $\|u\|=1$.) This equation allows us to answer the three questions stated previously.

1. Equation 12.6.7 is maximized when $\cos\theta=1$, i.e., when the gradient and $\vec{u}$ have the same direction. We conclude the gradient points in the direction of greatest z change.
2. Equation 12.6.7 is minimized when $\cos\theta=-1$, i.e., when the gradient and $\vec{u}$ have opposite directions. We conclude the gradient points in the opposite direction of the least z change.

3. Equation 12.6.7 is 0 when $\cos\theta = 0$, i.e., when the gradient and $\vec{u}$ are orthogonal to each other. We conclude the gradient is orthogonal to directions of no z change.

This result is rather amazing. Once again imagine standing in a rolling meadow and face the direction that leads you steepest uphill. Then the direction that leads steepest downhill is directly behind you, and side--stepping either left or right (i.e., moving perpendicularly to the direction you face) does not change your elevation at all.

Recall that a level curve is defined by a path in the xy-plane along which the z-values of a function do not change; the directional derivative in the direction of a level curve is 0. This is analogous to walking along a path in the rolling meadow along which the elevation does not change. The gradient at a point is orthogonal to the direction where the z does not change; i.e., the gradient is orthogonal to level curves.

Recall that a level curve is defined as a curve in the xy-plane along which the z-values of a function do not change. Let a surface $z = f(x,y)$ be given, and let's represent one such level curve as a vector--valued function, $\vec{r}(t) = \langle x(t), y(t)\rangle$. As the output of f does not change along this curve, $f\big(x(t), y(t)\big) = c$ for all t, for some constant c.

Since f is constant for all t, $\frac{df}{dt} = 0$. By the Multivariable Chain Rule, we also know

$$\begin{aligned}\frac{df}{dt} &= f_x(x,y)x'(t) + f_y(x,y)y'(t)\\ &= \langle f_x(x,y), f_y(x,y)\rangle \cdot \langle x'(t), y'(t)\rangle\\ &= \nabla f \cdot \vec{r}\,'(t)\\ &= 0.\end{aligned}$$

This last equality states $\nabla f \cdot \vec{r}\,'(t) = 0$: the gradient is orthogonal to the derivative of $\vec{r}$, meaning the gradient is orthogonal to $\vec{r}$ itself. Our conclusion: at any point on a surface, the gradient at that point is orthogonal to the level curve that passes through that point.

We restate these ideas in a theorem, then use them in an example.

theorem 111 The Gradient and Directional Derivatives

Let $z = f(x,y)$ be differentiable on an open set S with gradient ∇f, let $P = (x_0, y_0)$ be a point in S and let $\vec{u}$ be a unit vector.

1. The maximum value of $D_{\vec{u}}\, f(x_0, y_0)$ is $\|\nabla f(x_0, y_0)\|$; the direction of maximal z increase is $\nabla f(x_0, y_0)$.
2. The minimum value of $D_{\vec{u}}\, f(x_0, y_0)$ is $-\|\nabla f(x_0, y_0)\|$; the direction of minimal z increase is $-\nabla f(x_0, y_0)$.
3. At P, $\nabla f(x_0, y_0)$ is orthogonal to the level curve passing through $\big(x_0, y_0, f(x_0, y_0)\big)$.

Example 12.6.2: Finding directions of maximal and minimal increase

Let $f(x,y) = \sin x \cos y$ and let $P = (\pi/3, \pi/3)$. Find the directions of maximal/minimal increase, and find a direction where the instantaneous rate of z change is 0.

Solution

We begin by finding the gradient. $f_x = \cos x \cos y$ and $f_y = -\sin x \sin y$, thus

$$\nabla f = \langle \cos x \cos y, -\sin x \sin y\rangle \quad \text{and, at } P, \quad \nabla f\left(\frac{\pi}{3}, \frac{\pi}{3}\right) = \langle \frac{1}{4}, -\frac{3}{4}\rangle. \tag{12.6.8}$$

Thus the direction of maximal increase is $\langle 1/4, -3/4\rangle$. In this direction, the instantaneous rate of z change is $||\langle 1/4, -3/4\rangle|| = \sqrt{10}/4 \approx 0.79.$

Figure 12.17 shows the surface plotted from two different perspectives. In each, the gradient is drawn at P with a dashed line (because of the nature of this surface, the gradient points "into" the surface). Let $\vec{u} = \langle u_1, u_2\rangle$ be the unit vector in the direction of ∇f at P. Each graph of the figure also contains the vector $\langle u_1, u_2, ||\nabla f\,||\rangle$. This vector has a "run" of 1 (because in the xy-plane it moves 1 unit) and a "rise" of $||\nabla f\,||$, hence we can think of it as a vector with slope of $||\nabla f\,||$ in the direction of ∇f, helping us visualize how "steep" the surface is in its steepest direction.

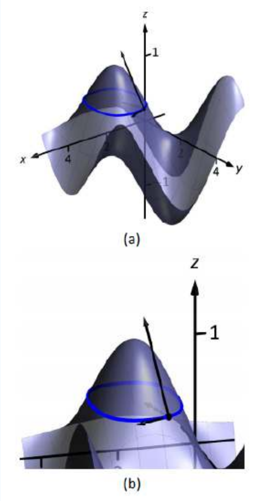


Figure 12.17: Graphing the surface and important directions in Example 12.6.2.

The direction of minimal increase is $\langle -1/4, 3/4\rangle$; in this direction the instantaneous rate of z change is $-\sqrt{10}/4 \approx -0.79$.

Any direction orthogonal to ∇f is a direction of no z change. We have two choices: the direction of $\langle 3,1\rangle$ and the direction of $\langle -3,-1\rangle$. The unit vector in the direction of $\langle 3,1\rangle$ is shown in each graph of the figure as well. The level curve at $z=\sqrt{3}/4$ is drawn: recall that along this curve the z-values do not change. Since $\langle 3,1\rangle$ is a direction of no z-change, this vector is tangent to the level curve at P.

Example 12.6.3: Understanding when $\nabla f = \vec{0}$

Let $f(x,y) = -x^2+2x-y^2+2y+1$. Find the directional derivative of f in any direction at $P=(1,1)$.

Solution

We find $\nabla f = \langle -2x+2, -2y+2\rangle$. At P, we have $\nabla f(1,1) = \langle 0,0\rangle$.

According to Theorem 111, this is the direction of maximal increase. However, $\langle 0,0\rangle$ is directionless; it has no displacement. And regardless of the unit vector $\vec{u}$ chosen, $D_{\vec{u}}\,f=0$.

Figure 12.18 helps us understand what this means. We can see that P lies at the top of a paraboloid. In all directions, the instantaneous rate of change is 0.

So what is the direction of maximal increase? It is fine to give an answer of $\vec{0}=\langle 0,0\rangle$, as this indicates that all directional derivatives are 0.

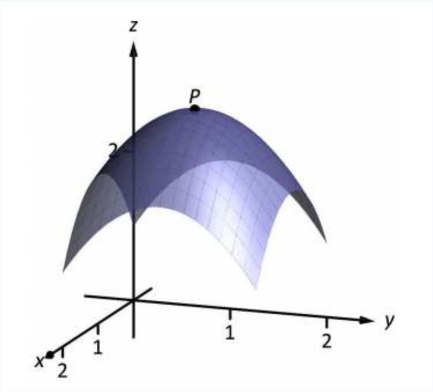


Figure 12.18: At the top of the paraboloid all directional derivatives are 0.

The fact that the gradient of a surface always points in the direction of steepest increase/decrease is very useful, as illustrated in the following example.

Example 12.6.4: The flow of water downhill

Consider the surface given by $f(x,y) = 20 - x^2 - 2y^2$. Water is poured on the surface at $(1, 1/4)$. What path does it take as it flows downhill?

Solution

Let $\vec{r}(t) = \langle x(t), y(t)\rangle$ be the vector--valued function describing the path of the water in the xy-plane; we seek $x(t)$ and $y(t)$. We know that water will always flow downhill in the steepest direction; therefore, at any point on its path, it will be moving in the direction of $-\nabla f$. (We ignore the physical effects of momentum on the water.) Thus $\vec{r}\,'(t)$ will be parallel to ∇f, and there is some constant c such that $c\nabla f = \vec{r}\,'(t) = \langle x'(t), y'(t)\rangle$.

We find $\nabla f = \langle -2x, -4y\rangle$ and write $x'(t)$ as $\frac{dx}{dt}$ and $y'(t)$ as $\frac{dy}{dt}$. Then

$$\begin{aligned} c\nabla f &= \langle x'(t), y'(t)\rangle \\ \langle -2cx, -4cy\rangle &= \langle \frac{dx}{dt}, \frac{dy}{dt}\rangle. \end{aligned}$$

This implies

$$-2cx = \frac{dx}{dt} \quad \text{and} \quad -4cy = \frac{dy}{dt}, \text{ i.e.,} \tag{12.6.9}$$

$$c = -\frac{1}{2x}\frac{dx}{dt} \quad \text{and} \quad c = -\frac{1}{4y}\frac{dy}{dt}. \tag{12.6.10}$$

As c equals both expressions, we have

$$\frac{1}{2x}\frac{dx}{dt} = \frac{1}{4y}\frac{dy}{dt}. \tag{12.6.11}$$

To find an explicit relationship between x and y, we can integrate both sides with respect to t. Recall from our study of differentials that $\frac{dx}{dt}dt = dx$. Thus:

$$\begin{aligned} \int \frac{1}{2x}\frac{dx}{dt}\,dt &= \int \frac{1}{4y}\frac{dy}{dt}\,dt \\ \int \frac{1}{2x}\,dx &= \int \frac{1}{4y}\,dy \\ \frac{1}{2}\ln|x| &= \frac{1}{4}\ln|y| + C_1 \\ 2\ln|x| &= \ln|y| + C_1 \\ \ln|x^2| &= \ln|y| + C_1 \end{aligned}$$

Now raise both sides as a power of e:

$$\begin{aligned} x^2 &= e^{\ln|y| + C_1} \\ x^2 &= e^{\ln|y|}e^{C_1} \qquad \text{(Note that } e^{C_1} \text{ is just a constant.)} \\ x^2 &= yC_2 \\ \frac{1}{C_2}x^2 &= y \qquad \text{(Note that } 1/C_2 \text{ is just a constant.)} \\ Cx^2 &= y. \end{aligned}$$

As the water started at the point $(1, 1/4)$, we can solve for C:

$$C(1)^2 = \frac{1}{4} \quad \Rightarrow \quad C = \frac{1}{4}. \tag{12.6.12}$$

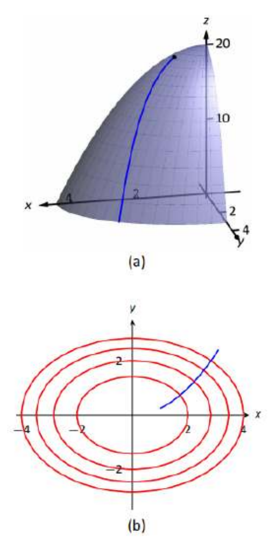


Figure 12.19: A graph of the surface described in Example 16.2.4 along with the path in the xy-plane with the level curves.

Thus the water follows the curve $y = x^2/4$ in the xy-plane. The surface and the path of the water is graphed in Figure 12.19(a). In part (b) of the figure, the level curves of the surface are plotted in the xy-plane, along with the curve $y = x^2/4$. Notice how the path intersects the level curves at right angles. As the path follows the gradient downhill, this reinforces the fact that the gradient is orthogonal to level curves.

Functions of Three Variables

The concepts of directional derivatives and the gradient are easily extended to three (and more) variables. We combine the concepts behind Definitions 90 and 91 and Theorem 110 into one set of definitions.

Definition 92 Directional Derivatives and Gradient with Three Variables

Let $w = F(x, y, z)$ be differentiable on an open ball B and let $\vec{u}$ be a unit vector in $\mathbb{R}^3$.

1. The **gradient** of F is $\nabla F = \langle F_x, F_y, F_z \rangle$.
2. The **directional derivative of** F **in the direction of** $\vec{u}$ is

$$D_{\vec{u}}\, F = \nabla F \cdot \vec{u}. \tag{12.6.13}$$

The same properties of the gradient given in Theorem 111, when f is a function of two variables, hold for F, a function of three variables.

THEOREM 112 The Gradient and Directional Derivatives with Three Variables

Let $w = F(x, y, z)$ be differentiable on an open ball B, let ∇F be the gradient of F, and let $\vec{u}$ be a unit vector.

1. The maximum value of $D_{\vec{u}}\, F$ is $\|\nabla F\|$, obtained when the angle between ∇F and $\vec{u}$ is 0, i.e., the direction of maximal increase is ∇F.
2. The minimum value of $D_{\vec{u}}\, F$ is $-\|\nabla F\|$, obtained when the angle between ∇F and $\vec{u}$ is π, i.e., the direction of minimal increase is $-\nabla F$.
3. $D_{\vec{u}}\, F = 0$ when ∇F and $\vec{u}$ are orthogonal.

We interpret the third statement of the theorem as "the gradient is orthogonal to level surfaces," the three--variable analogue to level curves.

Example 12.6.5: Finding directional derivatives with functions of three variables

If a point source S is radiating energy, the intensity I at a given point P in space is inversely proportional to the square of the distance between S and P. That is, when $S = (0, 0, 0)$, $I(x, y, z) = \frac{k}{x^2+y^2+z^2}$ for some constant k.

Let $k=1$, let $\vec{u}=\langle 2/3,2/3,1/3\rangle$ be a unit vector, and let $P=(2,5,3)$. Measure distances in inches. Find the directional derivative of I at P in the direction of $\vec{u}$, and find the direction of greatest intensity increase at P.

Solution
We need the gradient ∇I, meaning we need I_x, I_y and I_z. Each partial derivative requires a simple application of the Quotient Rule, giving

$$\begin{aligned}\nabla I &= \langle \frac{-2x}{(x^2+y^2+z^2)^2}, \frac{-2y}{(x^2+y^2+z^2)^2}, \frac{-2z}{(x^2+y^2+z^2)^2}\rangle \\ \nabla I(2,5,3) &= \langle \frac{-4}{1444}, \frac{-10}{1444}, \frac{-6}{1444}\rangle \approx \langle -0.003, -0.007, -0.004\rangle \\ D_{\vec{u}}\, I &= \nabla I(2,5,3)\cdot\vec{u} \\ &= -\frac{17}{2166} \approx -0.0078.\end{aligned}$$

The directional derivative tells us that moving in the direction of $\vec{u}$ from P results in a decrease in intensity of about -0.008 units per inch. (The intensity is decreasing as $\vec{u}$ moves one farther from the origin than P.)

The gradient gives the direction of greatest intensity increase. Notice that

$$\begin{aligned}\nabla I(2,5,3) &= \langle \frac{-4}{1444}, \frac{-10}{1444}, \frac{-6}{1444}\rangle \\ &= \frac{2}{1444}\langle -2,-5,-3\rangle.\end{aligned}$$

That is, the gradient at $(2,5,3)$ is pointing in the direction of $\langle -2,-5,-3\rangle$, that is, towards the origin. That should make intuitive sense: the greatest increase in intensity is found by moving towards to source of the energy.

The directional derivative allows us to find the instantaneous rate of z change in any direction at a point. We can use these instantaneous rates of change to define lines and planes that are *tangent* to a surface at a point, which is the topic of the next section.

This page titled 12.6: Directional Derivatives is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

12.7: Tangent Lines, Normal Lines, and Tangent Planes

Derivatives and tangent lines go hand-in-hand. Given $y=f(x)$, the line tangent to the graph of f at $x=x_0$ is the line through $\big(x_0,f(x_0)\big)$ with slope $f'(x_0)$; that is, the slope of the tangent line is the instantaneous rate of change of f at x_0.

When dealing with functions of two variables, the graph is no longer a curve but a surface. At a given point on the surface, it seems there are many lines that fit our intuition of being "tangent" to the surface.

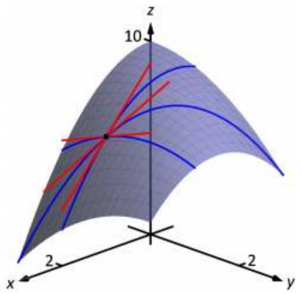


Figure 12.20: Showing various lines tangent to a surface.

In Figures 12.20 we see lines that are tangent to curves in space. Since each curve lies on a surface, it makes sense to say that the lines are also tangent to the surface. The next definition formally defines what it means to be "tangent to a surface."

Definition 93 Directional Tangent Line

Let $z=f(x,y)$ be differentiable on an open set S containing (x_0,y_0) and let $\vec{u}=\langle u_1,u_2\rangle$ be a unit vector.

1. The line ℓ_x through $\big(x_0,y_0,f(x_0,y_0)\big)$ parallel to $\langle 1,0,f_x(x_0,y_0)\rangle$ is the **tangent line to f in the direction of x at** (x_0,y_0).
2. The line ℓ_y through $\big(x_0,y_0,f(x_0,y_0)\big)$ parallel to $\langle 0,1,f_y(x_0,y_0)\rangle$ is the **tangent line to f in the direction of y at** (x_0,y_0).
3. The line $\ell_{\vec{u}}$ through $\big(x_0,y_0,f(x_0,y_0)\big)$ parallel to $\langle u_1,u_2,D_{\vec{u}}\,f(x_0,y_0)\rangle$ is the **tangent line to f in the direction of $\vec{u}$ at** (x_0,y_0).

It is instructive to consider each of three directions given in the definition in terms of "slope." The direction of ℓ_x is $\langle 1,0,f_x(x_0,y_0)\rangle$; that is, the "run" is one unit in the x-direction and the "rise" is $f_x(x_0,y_0)$ units in the z-direction. Note how the slope is just the partial derivative with respect to x. A similar statement can be made for ℓ_y. The direction of $\ell_{\vec{u}}$ is $\langle u_1,u_2,D_{\vec{u}}\,f(x_0,y_0)\rangle$; the "run" is one unit in the $\vec{u}$ direction (where $\vec{u}$ is a unit vector) and the "rise" is the directional derivative of z in that direction.

Definition 93 leads to the following parametric equations of directional tangent lines:

$$\ell_x(t)=\begin{cases}x=x_0+t\\ y=y_0\\ z=z_0+f_x(x_0,y_0)t\end{cases},\quad \ell_y(t)=\begin{cases}x=x_0\\ y=y_0+t\\ z=z_0+f_y(x_0,y_0)t\end{cases}\quad\text{and}\tag{12.7.1}$$

$$\ell_{\vec{u}}(t)=\begin{cases}x=x_0+u_1t\\ y=y_0+u_2t\\ z=z_0+D_{\vec{u}}\,f(x_0,y_0)t\end{cases}.$$

Example 12.7.1: Finding directional tangent lines

Find the lines tangent to the surface $z=\sin x\cos y$ at $(\pi/2,\pi/2)$ in the x and y directions and also in the direction of $\vec{v}=\langle -1,1\rangle$.

Solution

The partial derivatives with respect to x and y are:

$$\begin{aligned}f_x(x,y)&=\cos x\cos y &\Rightarrow\quad& f_x(\pi/2,\pi/2)=0\\ f_y(x,y)&=-\sin x\sin y &\Rightarrow\quad& f_y(\pi/2,\pi/2)=-1.\end{aligned}$$

At $(\pi/2,\pi/2)$, the z-value is 0.

Thus the parametric equations of the line tangent to f at $(\pi/2,\pi/2)$ in the directions of x and y are:

$$\ell_x(t)=\begin{cases}x=\pi/2+t\\ y=\pi/2\end{cases}\quad\text{and}\quad \ell_y(t)=\begin{cases}x=\pi/2\\ y=\pi/2+t\end{cases}.\tag{12.7.2}$$

$$z = 0 \qquad\qquad z = -t$$

The two lines are shown with the surface in Figure 12.21(a).

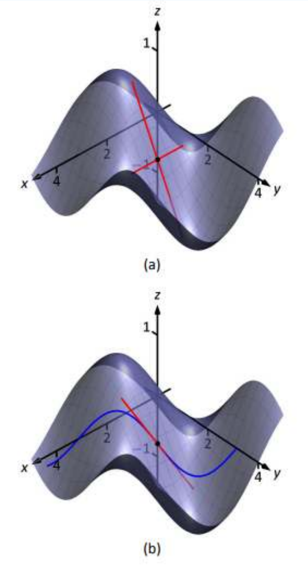


Figure 12.21: A surface and directional tangent lines in Example 12.7.1

To find the equation of the tangent line in the direction of $\vec{v}$, we first find the unit vector in the direction of $\vec{v}$: $\vec{u} = \langle -1/\sqrt{2}, 1/\sqrt{2}\rangle$. The directional derivative at $(\pi/2, \pi, 2)$ in the direction of $\vec{u}$ is

$$D_{\vec{u}}\, f(\pi/2, \pi, 2) = \langle 0, -1\rangle \cdot \langle -1/\sqrt{2}, 1/\sqrt{2}\rangle = -1/\sqrt{2}. \tag{12.7.3}$$

Thus the directional tangent line is

$$\ell_{\vec{u}}(t) = \begin{cases} x = \pi/2 - t/\sqrt{2} \\ y = \pi/2 + t/\sqrt{2} \\ z = -t/\sqrt{2} \end{cases}. \tag{12.7.4}$$

The curve through $(\pi/2, \pi/2, 0)$ in the direction of $\vec{v}$ is shown in Figure 12.21(b) along with $\ell_{\vec{u}}(t)$.

Example 12.7.2: Finding directional tangent lines

Let $f(x, y) = 4xy - x^4 - y^4$. Find the equations of *all* directional tangent lines to f at $(1, 1)$.

Solution

First note that $f(1, 1) = 2$. We need to compute directional derivatives, so we need ∇f. We begin by computing partial derivatives.

$$f_x = 4y - 4x^3 \Rightarrow f_x(1, 1) = 0; \quad f_y = 4x - 4y^3 \Rightarrow f_y(1, 1) = 0. \tag{12.7.5}$$

Thus $\nabla f(1, 1) = \langle 0, 0\rangle$. Let $\vec{u} = \langle u_1, u_2\rangle$ be any unit vector. The directional derivative of f at $(1, 1)$ will be $D_{\vec{u}}\, f(1, 1) = \langle 0, 0\rangle \cdot \langle u_1, u_2\rangle = 0$. It does not matter what direction we choose; the directional derivative is always 0. Therefore

$$\ell_{\vec{u}}(t) = \begin{cases} x = 1 + u_1 t \\ y = 1 + u_2 t \\ z = 2 \end{cases} \tag{12.7.6}$$

Figure 12.22 shows a graph of f and the point $(1, 1, 2)$. Note that this point comes at the top of a "hill," and therefore every tangent line through this point will have a "slope" of 0.

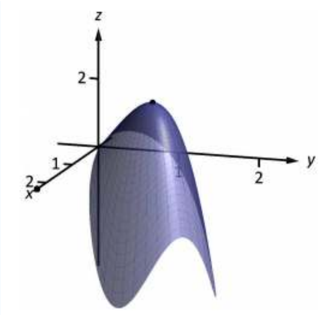


Figure 12.22: Graphing f in Example 12.7.2

That is, consider any curve on the surface that goes through this point. Each curve will have a relative maximum at this point, hence its tangent line will have a slope of 0. The following section investigates the points on surfaces where all tangent lines have a slope of 0.

Normal Lines

When dealing with a function $y=f(x)$ of one variable, we stated that a line through $(c,f(c))$ was *tangent* to f if the line had a slope of $f'(c)$ and was *normal* (or, *perpendicular, orthogonal*) to f if it had a slope of $-1/f'(c)$. We extend the concept of normal, or orthogonal, to functions of two variables.

Let $z=f(x,y)$ be a differentiable function of two variables. By Definition 93, at (x_0,y_0), $\ell_x(t)$ is a line parallel to the vector $\vec{d}_x = \langle 1,0,f_x(x_0,y_0)\rangle$ and $\ell_y(t)$ is a line parallel to $\vec{d}_y=\langle 0,1,f_y(x_0,y_0)\rangle$. Since lines in these directions through $\big(x_0,y_0,f(x_0,y_0)\big)$ are *tangent* to the surface, a line through this point and orthogonal to these directions would be *orthogonal*, or *normal*, to the surface. We can use this direction to create a normal line.

The direction of the normal line is orthogonal to $\vec{d}_x$ and $\vec{d}_y$, hence the direction is parallel to $\vec{d}_n=\vec{d}_x\times\vec{d}_y$. It turns out this cross product has a very simple form:

$$\vec{d}_x\times\vec{d}_y=\langle 1,0,f_x\rangle\times\langle 0,1,f_y\rangle=\langle -f_x,-f_y,1\rangle. \tag{12.7.7}$$

It is often more convenient to refer to the opposite of this direction, namely $\langle f_x,f_y,-1\rangle$. This leads to a definition.

Definition 94 Normal Line

Let $z=f(x,y)$ be differentiable on an open set S containing (x_0,y_0) where

$$a=f_x(x_0,y_0) \quad \text{and} \quad b=f_y(x_0,y_0) \tag{12.7.8}$$

are defined.

1. A nonzero vector parallel to $\vec{n}=\langle a,b,-1\rangle$ is **orthogonal to f at** $P=\big(x_0,y_0,f(x_0,y_0)\big)$.
2. The line ℓ_n through P with direction parallel to $\vec{n}$ is the **normal line to f at** P.

Thus the parametric equations of the normal line to a surface f at $\big(x_0,y_0,f(x_0,y_0)\big)$ is:

$$\ell_n(t)=\begin{cases} x=x_0+at\\ y=y_0+bt\\ z=f(x_0,y_0)-t\end{cases}. \tag{12.7.9}$$

Example 12.7.3: Finding a normal line

Find the equation of the normal line to $z=-x^2-y^2+2$ at $(0,1)$.

Solution

We find $z_x(x,y)=-2x$ and $z_y(x,y)=-2y$; at $(0,1)$, we have $z_x=0$ and $z_y=-2$. We take the direction of the normal line, following Definition 94, to be $\vec{n}=\langle 0,-2,-1\rangle$. The line with this direction going through the point $(0,1,1)$ is

$$\ell_n(t)=\begin{cases} x=0\\ y=-2t+1\end{cases} \quad \text{or} \quad \ell_n(t)=\langle 0,-2,-1\rangle t+\langle 0,1,1\rangle. \tag{12.7.10}$$

$$z = -t + 1$$

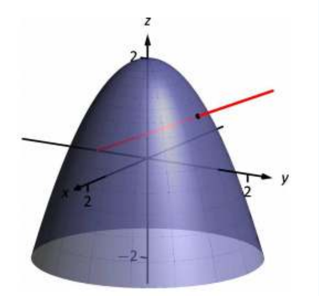


Figure 12.23: Graphing a surface with a normal line from Example 12.7.3

The surface $z = -x^2 + y^2$, along with the found normal line, is graphed in Figure 12.23.

The direction of the normal line has many uses, one of which is the definition of the **tangent plane** which we define shortly. Another use is in measuring distances from the surface to a point. Given a point Q in space, it is general geometric concept to define the distance from Q to the surface as being the length of the shortest line segment $\overline{PQ}$ over all points P on the surface. This, in turn, implies that $\vec{PQ}$ will be orthogonal to the surface at P. Therefore we can measure the distance from Q to the surface f by finding a point P on the surface such that $\vec{PQ}$ is parallel to the normal line to f at P.

Example 12.7.4: Finding the distance from a point to a surface

Let $f(x, y) = 2 - x^2 - y^2$ and let $Q = (2, 2, 2)$. Find the distance from Q to the surface defined by f.

Solution

This surface is used in Example 12.7.2, so we know that at (x, y), the direction of the normal line will be $\vec{d}_n = \langle -2x, -2y, -1\rangle$. A point P on the surface will have coordinates $(x, y, 2 - x^2 - y^2)$, so $\vec{PQ} = \langle 2 - x, 2 - y, x^2 + y^2\rangle$. To find where $\vec{PQ}$ is parallel to $\vec{d}_n$, we need to find x, y and c such that $c\vec{PQ} = \vec{d}_n$.

$$\begin{aligned} c\vec{PQ} &= \vec{d}_n \\ c\langle 2 - x, 2 - y, x^2 + y^2\rangle &= \langle -2x, -2y, -1\rangle. \end{aligned}$$

This implies

$$\begin{aligned} c(2 - x) &= -2x \\ c(2 - y) &= -2y \\ c(x^2 + y^2) &= -1 \end{aligned}$$

In each equation, we can solve for c:

$$c = \frac{-2x}{2 - x} = \frac{-2y}{2 - y} = \frac{-1}{x^2 + y^2}. \tag{12.7.11}$$

The first two fractions imply $x = y$, and so the last fraction can be rewritten as $c = -1/(2x^2)$. Then

$$\begin{aligned} \frac{-2x}{2 - x} &= \frac{-1}{2x^2} \\ -2x(2x^2) &= -1(2 - x) \\ 4x^3 &= 2 - x \\ 4x^3 + x - 2 &= 0. \end{aligned}$$

This last equation is a cubic, which is not difficult to solve with a numeric solver. We find that $x = 0.689$, hence $P = (0.689, 0.689, 1.051)$. We find the distance from Q to the surface of f is

$$\|\vec{PQ}\| = \sqrt{(2 - 0.689)^2 + (2 - 0.689)^2 + (2 - 1.051)^2} = 2.083. \tag{12.7.12}$$

We can take the concept of measuring the distance from a point to a surface to find a point Q a particular distance from a surface at a given point P on the surface.

Example $12.7.5$: Finding a point a set distance from a surface

Let $f(x,y) = x - y^2 + 3$. Let $P = \big(2, 1, f(2,1)\big) = (2,1,4)$. Find points Q in space that are 4 units from the surface of f at P. That is, find Q such that $\|\vec{PQ}\| = 4$ and $\vec{PQ}$ is orthogonal to f at P.

Solution

We begin by finding partial derivatives:

$$\begin{aligned} f_x(x,y) &= 1 & \Rightarrow \qquad & f_x(2,1) = 1 \\ f_y(x,y) &= -2y & \Rightarrow \qquad & f_y(2,1) = -2 \end{aligned}$$

The vector $\vec{n} = \langle 1, -2, -1\rangle$ is orthogonal to f at P. For reasons that will become more clear in a moment, we find the unit vector in the direction of $\vec{n}$:

$$\vec{u} = \frac{\vec{n}}{\|n\|} = \langle 1/\sqrt{6}, -2/\sqrt{6}, -1/\sqrt{6}\rangle \approx \langle 0.408, -0.816, -0.408\rangle. \tag{12.7.13}$$

Thus a the normal line to f at P can be written as

$$\ell_n(t) = \langle 2,1,4\rangle + t\langle 0.408, -0.816, -0.408\rangle. \tag{12.7.14}$$

An advantage of this parametrization of the line is that letting $t = t_0$ gives a point on the line that is $|t_0|$ units from P. (This is because the direction of the line is given in terms of a unit vector.) There are thus two points in space 4 units from P:

$$\begin{aligned} Q_1 &= \ell_n(4) & \qquad Q_2 &= \ell_n(-4) \\ &\approx \langle 3.63, -2.27, 2.37\rangle & &\approx \langle 0.37, 4.27, 5.63\rangle \end{aligned}$$

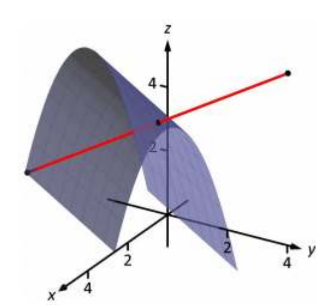


Figure 12.24: Graphing the surface in Example 12.7.5 along with points 4 units from the surface.

The surface is graphed along with points P, Q_1, Q_2 and a portion of the normal line to f at P.

Tangent Planes

We can use the direction of the normal line to define a plane. With $a = f_x(x_0,y_0)$, $b = f_y(x_0,y_0)$ and $P = \big(x_0, y_0, f(x_0,y_0)\big)$, the vector $\vec{n} = \langle a, b, -1\rangle$ is orthogonal to f at P. The plane through P with normal vector $\vec{n}$ is therefore **tangent** to f at P.

Definition 95 Tangent Plane

Let $z = f(x,y)$ be differentiable on an open set S containing (x_0,y_0), where $a = f_x(x_0,y_0)$, $b = f_y(x_0,y_0)$, $\vec{n} = \langle a, b, -1\rangle$ and $P = \big(x_0, y_0, f(x_0,y_0)\big)$.

The plane through P with normal vector $\vec{n}$ is the **tangent plane to** f **at** P. The standard form of this plane is

$$a(x - x_0) + b(y - y_0) - \big(z - f(x_0,y_0)\big) = 0. \tag{12.7.15}$$

Example 12.7.6: Finding tangent planes

Find the equation of the tangent plane to $z = -x^2 - y^2 + 2$ at $(0,1)$.

Solution

Note that this is the same surface and point used in Example 12.7.3. There we found $\vec{n} = \langle 0, -2, -1\rangle$ and $P = (0,1,1)$. Therefore the equation of the tangent plane is

$$-2(y-1) - (z-1) = 0. \tag{12.7.16}$$

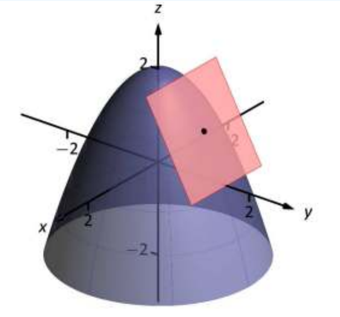


Figure 12.25: Graphing a surface with tangent plane from Example 17.2.6

The surface $z = -x^2 + y^2$ and tangent plane are graphed in Figure 12.25.

Example 12.7.7: Using the tangent plane to approximate function values

The point $(3, -1, 4)$ lies on the surface of an unknown differentiable function f where $f_x(3,-1) = 2$ and $f_y(3,-1) = -1/2$. Find the equation of the tangent plane to f at P, and use this to approximate the value of $f(2.9, -0.8)$.

Solution

Knowing the partial derivatives at $(3,-1)$ allows us to form the normal vector to the tangent plane, $\vec{n} = \langle 2, -1/2, -1\rangle$. Thus the equation of the tangent line to f at P is:

$$2(x-3) - 1/2(y+1) - (z-4) = 0 \quad \Rightarrow \quad z = 2(x-3) - 1/2(y+1) + 4. \tag{12.7.17}$$

Just as tangent lines provide excellent approximations of curves near their point of intersection, tangent planes provide excellent approximations of surfaces near their point of intersection. So $f(2.9, -0.8) \approx z(2.9, -0.8) = 3.7$.

This is not a new method of approximation. Compare the right hand expression for z in Equation 12.7.17 to the total differential:

$$dz = f_x dx + f_y dy \quad \text{and} \quad z = \underbrace{\underbrace{2}_{f_x}\underbrace{(x-3)}_{dx} + \underbrace{-1/2}_{f_y}\underbrace{(y+1)}_{dy}}_{dz} + 4. \tag{12.7.18}$$

Thus the "new z-value" is the sum of the change in z (i.e., dz) and the old z-value (4). As mentioned when studying the total differential, it is not uncommon to know partial derivative information about a unknown function, and tangent planes are used to give accurate approximations of the function.

The Gradient and Normal Lines, Tangent Planes

The methods developed in this section so far give a straightforward method of finding equations of normal lines and tangent planes for surfaces with explicit equations of the form $z = f(x,y)$. However, they do not handle implicit equations well, such as $x^2 + y^2 + z^2 = 1$. There is a technique that allows us to find vectors orthogonal to these surfaces based on the **gradient**.

Definition 96 Gradient

Let $w = F(x,y,z)$ be differentiable on an open ball B that contains the point (x_0, y_0, z_0).

1. The **gradient of** F is $\nabla F(x,y,z) = \langle f_x(x,y,z), f_y(x,y,z), f_z(x,y,z)\rangle$.
2. The **gradient of** F **at** (x_0, y_0, z_0) is

$$\nabla F(x_0,y_0,z_0) = \langle f_x(x_0,y_0,z_0), f_y(x_0,y_0,z_0), f_z(x_0,y_0,z_0)\rangle. \tag{12.7.19}$$

Recall that when $z=f(x,y)$, the gradient $\nabla f = \langle f_x, f_y\rangle$ is orthogonal to level curves of f. An analogous statement can be made about the gradient ∇F, where $w=F(x,y,z)$. Given a point (x_0,y_0,z_0), let $c=F(x_0,y_0,z_0)$. Then $F(x,y,z)=c$ is a **level surface** that contains the point (x_0,y_0,z_0). The following theorem states that $\nabla F(x_0,y_0,z_0)$ is orthogonal to this level surface.

THEOREM 113 The Gradient and Level Surfaces

Let $w=F(x,y,z)$ be differentiable on an open ball B containing (x_0,y_0,z_0) with gradient ∇F, where $F(x_0,y_0,z_0)=c$.

The vector $\nabla F(x_0,y_0,z_0)$ is orthogonal to the level surface $F(x,y,z)=c$ at (x_0,y_0,z_0).

The gradient at a point gives a vector orthogonal to the surface at that point. This direction can be used to find tangent planes and normal lines.

Example 12.7.8: Using the gradient to find a tangent plane

Find the equation of the plane tangent to the ellipsoid $\frac{x^2}{12}+\frac{y^2}{6}+\frac{z^2}{4}=1$ at $P=(1,2,1)$.

Solution

We consider the equation of the ellipsoid as a level surface of a function F of three variables, where $F(x,y,z)=\frac{x^2}{12}+\frac{y^2}{6}+\frac{z^2}{4}$. The gradient is:

$$\begin{aligned}\nabla F(x,y,z) &= \langle F_x, F_y, F_z\rangle \\ &= \langle \frac{x}{6}, \frac{y}{3}, \frac{z}{2}\rangle.\end{aligned}$$

At P, the gradient is $\nabla F(1,2,1)=\langle 1/6, 2/3, 1/2\rangle$. Thus the equation of the plane tangent to the ellipsoid at P is

$$\frac{1}{6}(x-1)+\frac{2}{3}(y-2)+\frac{1}{2}(z-1)=0. \tag{12.7.20}$$

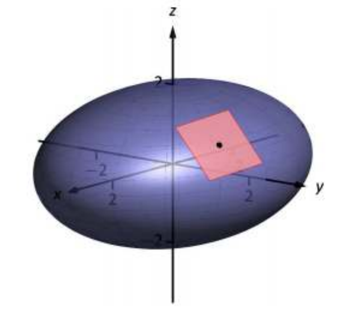


Figure 12.26: An ellipsoid and its tangent plane at a point.

The ellipsoid and tangent plane are graphed in Figure 12.26.

Tangent lines and planes to surfaces have many uses, including the study of instantaneous rates of changes and making approximations. Normal lines also have many uses. In this section we focused on using them to measure distances from a surface. Another interesting application is in computer graphics, where the effects of light on a surface are determined using normal vectors.

The next section investigates another use of partial derivatives: determining relative extrema. When dealing with functions of the form $y=f(x)$, we found relative extrema by finding x where $f'(x)=0$. We can *start* finding relative extrema of $z=f(x,y)$ by setting f_x and f_y to 0, but it turns out that there is more to consider.

12.7: Tangent Lines, Normal Lines, and Tangent Planes is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by LibreTexts.

12.8: Extreme Values

Given a function $z = f(x, y)$, we are often interested in points where z takes on the largest or smallest values. For instance, if z represents a cost function, we would likely want to know what (x, y) values minimize the cost. If z represents the ratio of a volume to surface area, we would likely want to know where z is greatest. This leads to the following definition.

Definition 97 Relative and Absolute Extrema

Let $z = f(x, y)$ be defined on a set S containing the point $P = (x_0, y_0)$.

1. If there is an open disk D containing P such that $f(x_0, y_0) \geq f(x, y)$ for all (x, y) in D, then f has a **relative maximum** at P; if $f(x_0, y_0) \leq f(x, y)$ for all (x, y) in D, then f has a **relative minimum** at P.
2. If $f(x_0, y_0) \geq f(x, y)$ for all (x, y) in S, then f has an **absolute maximum** at P; if $f(x_0, y_0) \leq f(x, y)$ for all (x, y) in S, then f has an **absolute minimum** at P.
3. If f has a relative maximum or minimum at P, then f has a **relative extrema** at P; if f has an absolute maximum or minimum at P, then f has an **absolute extrema** at P.

If f has a relative or absolute maximum at $P = (x_0, y_0)$, it means every curve on the surface of f through P will also have a relative or absolute maximum at P. Recalling what we learned in Section 3.1, the slopes of the tangent lines to these curves at P must be 0 or undefined. Since directional derivatives are computed using f_x and f_y, we are led to the following definition and theorem.

Definition 98 Critical Point

Let $z = f(x, y)$ be continuous on an open set S. A **critical point** $P = (x_0, y_0)$ of f is a point in S such that

- $f_x(x_0, y_0) = 0$ and $f_y(x_0, y_0) = 0$, or
- $f_x(x_0, y_0)$ and/or $f_y(x_0, y_0)$ is undefined.

theorem 114 Critical Points and Relative Extrema

Let $z = f(x, y)$ be defined on an open set S containing $P = (x_0, y_0)$. If f has a relative extrema at P, then P is a critical point of f.

Therefore, to find relative extrema, we find the critical points of f and determine which correspond to relative maxima, relative minima, or neither. The following examples demonstrate this process.

Example 12.8.1: Finding critical points and relative extrema

Let $f(x, y) = x^2 + y^2 - xy - x - 2$. Find the relative extrema of f.

Solution

We start by computing the partial derivatives of f:

$$f_x(x, y) = 2x - y - 1 \qquad \text{and} \qquad f_y(x, y) = 2y - x. \tag{12.8.1}$$

Each is never undefined. A critical point occurs when f_x and f_y are simultaneously 0, leading us to solve the following system of linear equations:

$$2x - y - 1 = 0 \qquad \text{and} \qquad -x + 2y = 0. \tag{12.8.2}$$

This solution to this system is $x = 2/3$, $y = 1/3$. (Check that at $(2/3, 1/3)$, both f_x and f_y are 0.)

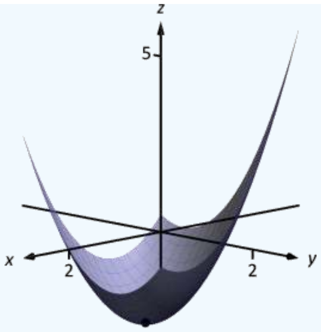


Figure 12.27: The surface in Example 12.8.1 with its absolute minimum indicated.

The graph in Figure 12.27 shows f along with this critical point. It is clear from the graph that this is a relative minimum; further consideration of the function shows that this is actually the absolute minimum.

Example $12.8.2$: Finding critical points and relative extrema

Let $f(x,y)=-\sqrt{x^2+y^2}+2$. Find the relative extrema of f.

Solution

We start by computing the partial derivatives of f:

$$f_x(x,y)=\frac{-x}{\sqrt{x^2+y^2}} \qquad \text{and} \qquad f_y(x,y)=\frac{-y}{\sqrt{x^2+y^2}}. \tag{12.8.3}$$

It is clear that $f_x=0$ when $x=0$ \& $y\neq 0$, and that $f_y=0$ when $y=0$ \& $x\neq 0$. At $(0,0)$, both f_x and f_y are *not* 0, but rather undefined. The point $(0,0)$ is still a critical point, though, because the partial derivatives are undefined. This is the only critical point of f.

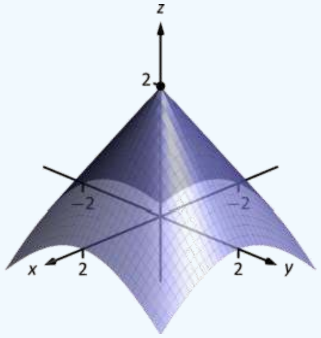


Figure 12.28: The surface in Example 12.8.2 with its absolute maximum indicated.

The surface of f is graphed in Figure 12.28 along with the point $(0,0,2)$. The graph shows that this point is the absolute maximum of f.

In each of the previous two examples, we found a critical point of f and then determined whether or not it was a relative (or absolute) maximum or minimum by graphing. It would be nice to be able to determine whether a critical point corresponded to a max or a min without a graph. Before we develop such a test, we do one more example that sheds more light on the issues our test needs to consider.

Example $12.8.3$: Finding critical points and relative extrema

Let $f(x,y)=x^3-3x-y^2+4y$. Find the relative extrema of f.

Solution

Once again we start by finding the partial derivatives of f:

$$f_x(x,y)=3x^2-3 \qquad \text{and} \qquad f_y(x,y)=-2y+4. \tag{12.8.4}$$

Each is always defined. Setting each equal to 0 and solving for x and y, we find

$$\begin{aligned} f_x(x,y) = 0 &\quad \Rightarrow x = \pm 1 \\ f_y(x,y) = 0 &\quad \Rightarrow y = 2. \end{aligned}$$

We have two critical points: $(-1,2)$ and $(1,2)$. To determine if they correspond to a relative maximum or minimum, we consider the graph of f in Figure 12.29.

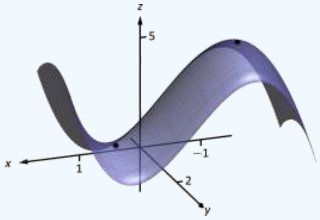


Figure 12.29: The surface in Example 12.8.3 with both critical points marked.

The critical point $(-1,2)$ clearly corresponds to a relative maximum. However, the critical point at $(1,2)$ is neither a maximum nor a minimum, displaying a different, interesting characteristic.

If one walks parallel to the y-axis towards this critical point, then this point becomes a relative maximum along this path. But if one walks towards this point parallel to the x-axis, this point becomes a relative minimum along this path. A point that seems to act as both a max and a min is a **saddle point**. A formal definition follows.

Definition 99 Saddle Point

Let $P = (x_0, y_0)$ be in the domain of f where $f_x = 0$ and $f_y = 0$ at P. P is a **saddle point** of f if, for every open disk D containing P, there are points (x_1, y_1) and (x_2, y_2) in D such that $f(x_0, y_0) > f(x_1, y_1)$ and $f(x_0, y_0) < f(x_2, y_2)$.

At a saddle point, the instantaneous rate of change in all directions is 0 and there are points nearby with z-values both less than and greater than the z-value of the saddle point.

Before Example 12.8.3 we mentioned the need for a test to differentiate between relative maxima and minima. We now recognize that our test also needs to account for saddle points. To do so, we consider the second partial derivatives of f.

Recall that with single variable functions, such as $y = f(x)$, if $f''(c) > 0$, then f is concave up at c, and if $f'(c) = 0$, then f has a relative minimum at $x = c$. (We called this the Second Derivative Test.) Note that at a saddle point, it seems the graph is "both" concave up and concave down, depending on which direction you are considering.

It would be nice if the following were true:

$$\begin{array}{rcc} f_{xx} \text{ and } f_{yy} > 0 & \Rightarrow & \text{relative minimum} \\ f_{xx} \text{ and } f_{yy} < 0 & \Rightarrow & \text{relative maximum} \\ f_{xx} \text{ and } f_{yy} \text{ have opposite signs} & \Rightarrow & \text{saddle point.} \end{array}$$

However, this is not the case. Functions f exist where f_{xx} and f_{yy} are both positive but a saddle point still exists. In such a case, while the concavity in the x-direction is up (i.e., $f_{xx} > 0$) and the concavity in the y-direction is also up (i.e., $f_{yy} > 0$), the concavity switches somewhere in between the x- and y-directions.

To account for this, consider $D = f_{xx}f_{yy} - f_{xy}f_{yx}$. Since f_{xy} and f_{yx} are equal when continuous (refer back to Theorem 103), we can rewrite this as $D = f_{xx}f_{yy} - f_{xy}^2$. D can be used to test whether the concavity at a point changes depending on direction. If $D > 0$, the concavity does not switch (i.e., at that point, the graph is concave up or down in all directions). If $D < 0$, the concavity does switch. If $D = 0$, our test fails to determine whether concavity switches or not. We state the use of D in the following theorem.

THEOREM 115 Second Derivative Test

Let $z = f(x,y)$ be differentiable on an open set containing $P = (x_0, y_0)$, and let

$$D = f_{xx}(x_0,y_0) f_{yy}(x_0,y_0) - f_{xy}^2(x_0,y_0). \tag{12.8.5}$$

1. If $D > 0$ and $f_{xx}(x_0, y_0) > 0$, then P is a relative minimum of f.

2. If $D > 0$ and $f_{xx}(x_0, y_0) < 0$, then P is a relative maximum of f.
3. If $D < 0$, then P is a saddle point of f.
4. If $D = 0$, the test is inconclusive.

We first practice using this test with the function in the previous example, where we visually determined we had a relative maximum and a saddle point.

Example 12.8.4: Using the Second Derivative Test

Let $f(x, y) = x^3 - 3x - y^2 + 4y$ as in Example 12.8.3. Determine whether the function has a relative minimum, maximum, or saddle point at each critical point.

Solution

We determined previously that the critical points of f are $(-1, 2)$ and $(1, 2)$. To use the Second Derivative Test, we must find the second partial derivatives of f:

$$f_{xx} = 6x; \qquad f_{yy} = -2; \qquad f_{xy} = 0. \tag{12.8.6}$$

Thus $D(x, y) = -12x$.

At $(-1, 2)$: $D(-1, 2) = 12 > 0$, and $f_{xx}(-1, 2) = -6$. By the Second Derivative Test, f has a relative maximum at $(-1, 2)$.

At $(1, 2)$: $D(1, 2) = -12 < 0$. The Second Derivative Test states that f has a saddle point at $(1, 2)$.

The Second Derivative Test confirmed what we determined visually.

Example 12.8.5: Using the Second Derivative Test

Find the relative extrema of $f(x, y) = x^2y + y^2 + xy$.

Solution

We start by finding the first and second partial derivatives of f:

$$\begin{aligned} f_x &= 2xy + y & f_y &= x^2 + 2y + x \\ f_{xx} &= 2y & f_{yy} &= 2 \\ f_{xy} &= 2x + 1 & f_{yx} &= 2x + 1. \end{aligned} \tag{12.8.7}$$

We find the critical points by finding where f_x and f_y are simultaneously 0 (they are both never undefined). Setting $f_x = 0$, we have:

$$f_x = 0 \quad \Rightarrow \quad 2xy + y = 0 \quad \Rightarrow \quad y(2x + 1) = 0. \tag{12.8.8}$$

This implies that for $f_x = 0$, either $y = 0$ or $2x + 1 = 0$.

Assume $y = 0$ then consider $f_y = 0$:

$$\begin{aligned} f_y &= 0 \\ x^2 + 2y + x &= 0, \qquad \text{and since } y = 0, \text{ we have} \\ x^2 + x &= 0 \\ x(x + 1) &= 0. \end{aligned}$$

Thus if $y = 0$, we have either $x = 0$ or $x = -1$, giving two critical points: $(-1, 0)$ and $(0, 0)$.

Going back to f_x, now assume $2x + 1 = 0$, i.e., that $x = -1/2$, then consider $f_y = 0$:

$$\begin{aligned} f_y &= 0 \\ x^2 + 2y + x &= 0, \qquad \text{and since } x = -1/2, \text{ we have} \\ 1/4 + 2y - 1/2 &= 0 \\ y &= 1/8. \end{aligned}$$

Thus if $x = -1/2$, $y = 1/8$ giving the critical point $(-1/2, 1/8)$.

With $D = 4y - (2x+1)^2$, we apply the Second Derivative Test to each critical point.

At $(-1, 0)$, $D < 0$, so $(-1, 0)$ is a saddle point.

At $(0, 0)$, $D < 0$, so $(0, 0)$ is also a saddle point.

At $(-1/2, 1/8)$, $D > 0$ and $f_{xx} > 0$, so $(-1/2, 1/8)$ is a relative minimum.

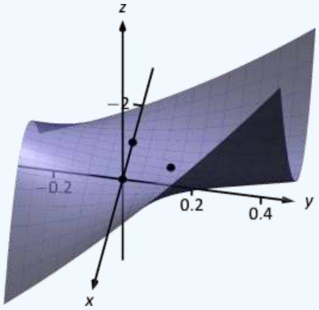


Figure 12.30: Graphing f from Example 12.8.5 and its relative extrema.

Figure 12.30 shows a graph of f and the three critical points. Note how this function does not vary much near the critical points -- that is, visually it is difficult to determine whether a point is a saddle point or relative minimum (or even a critical point at all!). This is one reason why the Second Derivative Test is so important to have.

Constrained Optimization

When optimizing functions of one variable such as $y = f(x)$, we made use of Theorem 25, the Extreme Value Theorem, that said that over a closed interval I, a continuous function has both a maximum and minimum value. To find these maximum and minimum values, we evaluated f at all critical points in the interval, as well as at the endpoints (the "boundary") of the interval.

A similar theorem and procedure applies to functions of two variables. A continuous function over a closed set also attains a maximum and minimum value (see the following theorem). We can find these values by evaluating the function at the critical values in the set and over the boundary of the set. After formally stating this extreme value theorem, we give examples.

theorem 116 Extreme Value Theorem

Let $z = f(x, y)$ be a continuous function on a closed, bounded set S. Then f has a maximum and minimum value on S.

Example $12.8.6$: Finding extrema on a closed set

Let $f(x, y) = x^2 - y^2 + 5$ and let S be the triangle with vertices $(-1, -2)$, $(0, 1)$ and $(2, -2)$. Find the maximum and minimum values of f on S.

Solution
It can help to see a graph of f along with the set S. In Figure 12.31(a) the triangle defining S is shown in the xy-plane in a dashed line. Above it is the surface of f; we are only concerned with the portion of f enclosed by the "triangle" on its surface.

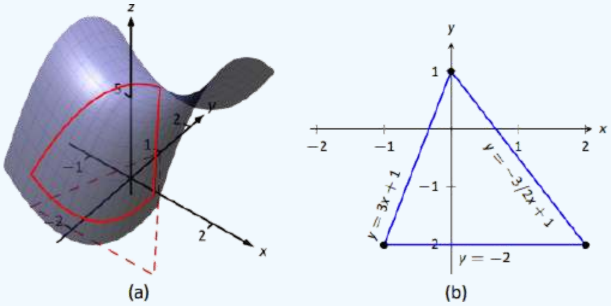


Figure 12.31: Plotting the surface of f along with the restricted domain S.

We begin by finding the critical points of f. With $f_x = 2x$ and $f_y = -2y$, we find only one critical point, at $(0, 0)$.

We now find the maximum and minimum values that f attains along the boundary of S, that is, along the edges of the triangle. In Figure 12.31(b) we see the triangle sketched in the plane with the equations of the lines forming its edges labeled.

Start with the bottom edge, along the line $y=-2$. If y is -2, then on the surface, we are considering points $f(x,-2)$; that is, our function reduces to $f(x,-2)=x^2-(-2)^2+5=x^2+1=f_1(x)$. We want to maximize/minimize $f_1(x)=x^2+1$ on the interval $[-1,2]$. To do so, we evaluate $f_1(x)$ at its critical points and at the endpoints.

The critical points of f_1 are found by setting its derivative equal to 0:

$$f_1'(x)=0 \qquad \Rightarrow x=0. \tag{12.8.9}$$

Evaluating f_1 at this critical point, and at the endpoints of $[-1,2]$ gives:

$$\begin{aligned} f_1(-1)=2 &\qquad\Rightarrow\qquad f(-1,-2)=2\\ f_1(0)=1 &\qquad\Rightarrow\qquad f(0,-2)=1\\ f_1(2)=5 &\qquad\Rightarrow\qquad f(2,-2)=5. \end{aligned}$$

Notice how evaluating f_1 at a point is the same as evaluating f at its corresponding point.

We need to do this process twice more, for the other two edges of the triangle.

Along the left edge, along the line $y=3x+1$, we substitute $3x+1$ in for y in $f(x,y)$:

$$f(x,y)=f(x,3x+1)=x^2-(3x+1)^2+5=-8x^2-6x+4=f_2(x). \tag{12.8.10}$$

We want the maximum and minimum values of f_2 on the interval $[-1,0]$, so we evaluate f_2 at its critical points and the endpoints of the interval. We find the critical points:

$$f_2'(x)=-16x-6=0 \qquad\Rightarrow\qquad x=-3/8. \tag{12.8.11}$$

Evaluate f_2 at its critical point and the endpoints of $[-1,0]$:

$$\begin{aligned} f_2(-1)=2 &\qquad\Rightarrow\qquad f(-1,-2)=2\\ f_2(-3/8)=41/8=5.125 &\qquad\Rightarrow\qquad f(-3/8,-0.125)=5.125\\ f_2(0)=1 &\qquad\Rightarrow\qquad f(0,1)=4. \end{aligned}$$

Finally, we evaluate f along the right edge of the triangle, where $y=-3/2x+1$.

$$\begin{aligned} f(x,y)=f(x,-3/2x+1)&=x^2-(-3/2x+1)^2+5\\ &=-\frac{5}{4}x^2+3x+4=f_3(x). \end{aligned} \tag{12.8.12}$$

The critical points of $f_3(x)$ are:

$$f_3'(x)=0 \qquad\Rightarrow\qquad x=6/5=1.2. \tag{12.8.13}$$

We evaluate f_3 at this critical point and at the endpoints of the interval $[0,2]$:

$$\begin{aligned} f_3(0)=4 &\qquad\Rightarrow\qquad f(0,1)=4\\ f_3(1.2)=5.8 &\qquad\Rightarrow\qquad f(1.2,-0.8)=5.8\\ f_3(2)=5 &\qquad\Rightarrow\qquad f(2,-2)=5. \end{aligned}$$

One last point to test: the critical point of f, $(0,0)$. We find $f(0,0)=5$.

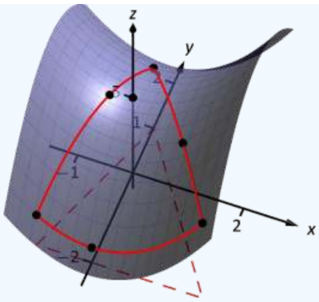


Figure 12.32: The surface of f along with important points along the boundary of S and the interior.

We have evaluated f at a total of 7 different places, all shown in Figure 12.32. We checked each vertex of the triangle twice, as each showed up as the endpoint of an interval twice. Of all the z-values found, the maximum is 5.8, found at $(1.2, -0.8)$; the minimum is 1, found at $(0, -2)$.

This portion of the text is entitled "Constrained Optimization" because we want to optimize a function (i.e., find its maximum and/or minimum values) subject to a *constraint* -- limits on which input points are considered. In the previous example, we constrained ourselves by considering a function only within the boundary of a triangle. This was largely arbitrary; the function and the boundary were chosen just as an example, with no real "meaning" behind the function or the chosen constraint.

However, solving constrained optimization problems is a very important topic in applied mathematics. The techniques developed here are the basis for solving larger problems, where more than two variables are involved.

We illustrate the technique once more with a classic problem.

Example $12.8.7$: Constrained Optimization

The U.S. Postal Service states that the girth+length of Standard Post Package must not exceed 130". Given a rectangular box, the "length" is the longest side, and the "girth" is twice the width+height.

Given a rectangular box where the width and height are equal, what are the dimensions of the box that give the maximum volume subject to the constraint of the size of a Standard Post Package?

Solution

Let w, h and ℓ denote the width, height and length of a rectangular box; we assume here that $w = h$. The girth is then $2(w+h)$ $= 4w$. The volume of the box is $V(w,\ell) = wh\ell = w^2\ell$. We wish to maximize this volume subject to the constraint $4w + \ell \leq 130$, or $\ell \leq 130 - 4w$. (Common sense also indicates that $\ell > 0, w > 0$.)

We begin by finding the critical values of V. We find that $V_w = 2w\ell$ and $V_\ell = w^2$; these are simultaneously 0 only at $(0,0)$. This gives a volume of 0, so we can ignore this critical point.

We now consider the volume along the constraint $\ell = 130 - 4w$. Along this line, we have:

$$V(w,\ell) = V(w, 130-4w) = w^2(130-4w) = 130w^2 - 4w^3 = V_1(w). \tag{12.8.14}$$

The constraint is applicable on the w-interval $[0, 32.5]$ as indicated in the figure. Thus we want to maximize V_1 on $[0, 32.5]$.

Finding the critical values of V_1, we take the derivative and set it equal to 0:

$$\begin{aligned} V\,'_1(w) = 260w - 12w^2 = 0 \quad &\Rightarrow \quad w(260-12w) \\ = 0 \quad \Rightarrow \quad w = 0, &\frac{260}{12} \approx 21.67. \end{aligned} \tag{12.8.15}$$

We found two critical values: when $w = 0$ and when $w = 21.67$. We again ignore the $w = 0$ solution; the maximum volume, subject to the constraint, comes at $w = h = 21.67$, $\ell = 130 - 4(21.6) = 43.33$. This gives a volume of $V(21.67, 43.33) \approx 19,408\text{in}^3$.

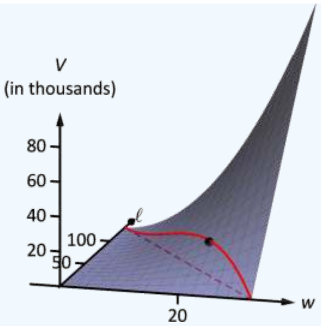


Figure 12.33: Graphing the volume of a box with girth 4w and length l, subject to a size constraint.

The volume function $V(w, \ell)$ is shown in Figure 12.33 along with the constraint $\ell = 130 - 4w$. As done previously, the constraint is drawn dashed in the xy-plane and also along the surface of the function. The point where the volume is maximized is indicated.

It is hard to overemphasize the importance of optimization. In "the real world," we routinely seek to make *something* better. By expressing the *something* as a mathematical function, "making *something* better" means "optimize *some function*."

The techniques shown here are only the beginning of an incredibly important field. Many functions that we seek to optimize are incredibly complex, making the step of "find the gradient and set it equal to $\vec{0}$" highly nontrivial. Mastery of the principles here is key to being able to tackle these more complicated problems.

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 12.8: Extreme Values is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

12.E: Applications of Functions of Several Variables (Exercises)

12.1: Introduction to Multivariable Functions

Terms and Concepts

1. Give two examples (other than those given in the text) of "real world" functions that require more than one input.

2. The graph of a function of two variables is a ________.

3. Most people are familiar with the concept of level curves in the context of ______ maps.

4. T/F: Along a level curve, the output of a function does not change.

5. The analogue of a level curve for functions of three variables is a level ________.

6. What does it mean when level curves are close together? Far apart?

Problems

In Exercises 7-14, give the domain and range of the multivariable function.

7. $f(x,y)=x^2+y^2+2$

8. $f(x,y)=x+2y$

9. $f(x,y)=x-2y$

10. $f(x,y)=\frac{1}{x+2y}$

11. $f(x,y)=\frac{1}{x^2+y^2+1}$

12. $f(x,y)=\sin x\cos y$

13. $f(x,y)=\sqrt{9-x^2-y^2}$

14. $f(x,y)=\frac{1}{\sqrt{x^2+y^2-9}}$

In Exercises 15-22, describe in words and sketch the level curves for the function and given *c* values.

15. $f(x,y)=3x-2y;\ c=-2,0,2$

16. $f(x,y)=x^2-y^2;\ c=-1,0,1$

17. $f(x,y)=x-y^2;\ c=-2,0,2$

18. $f(x,y)=\frac{1-x^2-y^2}{2y-2x};\ c=-2,0,2$

19. $f(x,y)=\frac{2x-2y}{x^2+y^2+1};\ c=-1,0,1$

20. $f(x,y)=\frac{y-x^3-1}{x};\ c=-3,-1,0,1,3$

21. $f(x,y)=\sqrt{x^2+4y^2};\ c=1,2,3,4$

22. $f(x,y)=x^2+4y^2;\ c=1,2,3,4$

In Exercises 23-26, give the domain and range of the functions of three variables.

23. $f(x,y,z)=\frac{x}{x+2y-4z}$

24. $f(x,y,z)=\frac{1}{1-x^2-y^2-z^2}$

25. $f(x,y,z)=\sqrt{z-x^2+y^2}$

26. $f(x,y,z)=z^2\sin x\cos y$

In Exercises 27-30, describe the level surfaces of the given functions of three variables.

27. $f(x,y,z)=x^2+y^2+z^2$

28. $f(x,y,z)=z-x^2+y^2$

29. $f(x,y,z)=\frac{x^2+y^2}{z}$

30. $f(x,y,z)=\frac{z}{x-y}$

31. Compare the level curves of Exercises 21 and 22. How are they similar, and how are they different? Each surface is a quadric surface; describe how the level curves are consistent with what we know about each surface.

12.2: Limits and Continuity of Multivariable Functions

Terms and Concepts

1. Describe in your own words the difference between boundary and interior point of a set.

2. Use your own words to describe (informally) what $\lim\limits_{(x,y)\to(1,2)} f(x,y)=17$ means.

3. Give an example of a closed, bounded set.

4. Give an example of a closed, unbounded set.

5. Give an example of a open, bounded set.

6. Give an example of a open, unbounded set.

Problems

In Exercises 7-10, a set *S* is given.
(a) Give one boundary point and one interior point, when possible, of *S*.
(b) State whether *S* is open, closed, or neither.
(c) State whether *S* is bounded or unbounded.

7. $S=\left\{(x,y)\middle|\frac{(x-1)^2}{4}+\frac{(y-3)^2}{9}\leq 1\right\}$

8. $S=\left\{(x,y)\middle|y\neq x^2\right\}$

9. $S=\left\{(x,y)\middle|x^2+y^2=1\right\}$

10. $S=\left\{(x,y)\middle|y>\sin x\right\}$

In Exercises 11-14:
(a) Find the domain *D* of the given function.
(b) State whether *D* is open or closed set.
(c) State whether *D* is bounded or unbounded.

11. $f(x,y)=\sqrt{9-x^2-y^2}$

12. $f(x,y)=\sqrt{y-x^2}$

13. $f(x,y)=\frac{1}{\sqrt{y-x^2}}$

14. $f(x,y)=\frac{x^2-y^2}{x^2+y^2}$

In Exercises 15-20, a limit is given. Evaluate the limit along the paths given, then state why these results show the given limit does not exist.

15. $\lim\limits_{(x,y)\to(0,0)}\frac{x^2-y^2}{x^2+y^2}$
(a) Along the path $y=0$.
(b) Along the path $x=0$.

16. $\lim\limits_{(x,y)\to(0,0)}\frac{x+y}{x-y}$
(a) Along the path $y=mx$.

17. $\lim\limits_{(x,y)\to(0,0)}\frac{xy-y^2}{y^2+x}$
(a) Along the path $y=mx$.
(b) Along the path $x=0$.

18. $\lim_{(x,y)\to(0,0)} \frac{\sin(x^2)}{y}$

(a) Along the path $y = mx$.
(b) Along the path $y = x^2$.

19. $\lim_{(x,y)\to(1,2)} \frac{x+y-3}{x^2-1}$

(a) Along the path $y = 2$.
(b) Along the path $y = x + 1$.

20. $\lim_{(x,y)\to(\pi,\pi/2)} \frac{\sin x}{\cos y}$

(a) Along the path $x = \pi$.
(b) Along the path $y = x - \pi/2$.

12.3: Partial Derivatives

Terms and Concepts

1. What is the difference between a constant and a coefficient?

2. Given a function $z = f(x,y)$, explain in your own words how to compute f_x.

3. In the mixed partial fraction f_{xy}, which is computed first, f_x or f_y?

4. In the mixed partial fraction $\frac{\partial^2 f}{\partial x \partial y}$, which is computed first, f_x or f_y.

Problems

In Exercises 5-8, evaluate $f_x(x,y)$ and $f_y(x,y)$ at the indicated point.

5. $f(x,y) = x^2y - x + 2y + 3 \text{ at } (1,2)$

6. $f(x,y) = x^3 - 3x + y^2 - 6y \text{ at } (-1,3)$

7. $f(x,y) = \sin y \cos x \text{ at } (\pi/3, \pi/3)$

8. $f(x,y) = \ln(xy) \text{ at } (-2,-3)$

In Exercises 9-26, find $f_x, f_y, f_{xx}, f_{yy}, f_{xy}$ and f_{yx}.

9. $f(x,y) = x^2y + 3x^2 + 4y - 5$

10. $f(x,y) = y^3 + 3xy^2 + 3x^2y + x^3$

11. $f(x,y) = \frac{x}{y}$

12. $f(x,y) = \frac{4}{xy}$

13. $f(x,y) = e^{x^2+y^2}$

14. $f(x,y) = e^{x+2y}$

15. $f(x,y) = \sin x \cos y$

16. $f(x,y) = (x+y)^3$

17. $f(x,y) = \cos(5xy^3)$

18. $f(x,y) = \sin(5x^2 + 2y^3)$

19. $f(x,y) = \sqrt{4xy^2 + 1}$

20. $f(x,y) = (2x + 5y)\sqrt{y}$

21. $f(x,y) = \frac{1}{x^2+y^2+1}$

22. $f(x,y) = 5x - 17y$

23. $f(x,y) = 3x^2 + 1$

24. $f(x,y) = \ln(x^2 + y)$

25. $f(x,y)=\frac{\ln x}{4y}$

26. $f(x,y)=5e^x\sin y+9$

In Exercises 27-30, form a function $z=f(x,y)$ such that f_x and f_y match those given.

27. $f_x=\sin y+1,\quad f_y=x\cos y$

28. $f_x=x+y,\quad f_y=x+y$

29. $f_x=6xy-4y^2,\quad f_y=3x^2-8xy+2$

30. $f_x=\frac{2x}{x^2+y^2},\quad f_y=\frac{2y}{x^2+y^2}$

In Exercises 31-34, find f_x, f_y, f_z, f_{yz} and f_{zy}.

31. $f(x,y,z)=x^2e^{2y-3z}$

32. $f(x,y,z)=x^3y^2+x^3z+y^2z$

33. $f(x,y,z)=\frac{3x}{7y^2z}$

34. $f(x,y,z)=\ln(xyz)$

12.4: Differentiability and the Total Differential

Terms and Concepts

1. T/F: If $f(x,y)$ is differentiable on *S*, the f is continuous on *S*.

2. T/F: If f_x and f_y are continuous on *S*, then f is differentiable on *S*.

3. T/F: If $z=f(x,y)$ is differentiable, then the change in *z* over small changes dx and dy in *x* and *y* is approximately *dz*.

4. Finish the sentence: "The new *z*-value is approximately the old *z*-value plus the approximate _______."

Problems

In Exercises 5-8, find the total differential *dz*.

5. $z=x\sin y+x^2$

6. $z=(2x^2+3y)^2$

7. $z=5x-7y$

8. $z=xe^{x+y}$

In Exercises 9-12, a function $z=f(x,y)$ is given. Give the indicated approximation using the total differential.

9. $f(x,y)=\sqrt{x^2+y}$. Approximate $f(2,95,7.1)$ knowing $f(3,7)=4$.

10. $f(x,y)=\sin x\cos y$. Approximate $f(0.1,-0.1)$ knowing $f(0,0)=0$.

11. $f(x,y)=x^2y-xy^2$. Approximate $f(2.04,3.06)$ knowing $f(2,3)=-6$.

12. $f(x,y)=\ln(x-y)$. Approximate $f(5.1,3.98)$ knowing $f(5,4)=0$.

Exercises 13-16 ask a variety of questions dealing with approximating error and sensitivity analysis.

13. A cylindrical storage tank is to be 2ft tall with a radius of 1ft. Is the volume of the tank more sensitive to changes in the radius or the height?

14. **Projectile Motion:** The *x*-value of an object moving under the principles of projectile motion is $x(\theta,v_o,t)=(v_o\cos\theta)t$. A particular projectile is fired with an initial velocity of $v_o=250$ft/s and an angle of elevation of $\theta=60^\circ$. It travels a distance of 375ft in 3 seconds. Is the projectile more sensitive to errors in initial speed or angle of elevations?

15. The length l of a long wall is to be approximated. The angle θ, as shown in the diagram (not to scale), is measured to be 85°, and the distance *x* is measured to be $30'$. Assume that the triangle formed is a right triangle.

Is the measurement of the length of l more sensitive to errors in the measurement of *x* or $\ln\theta$.

16. It is "common sense" that it is far better to measure a long distance with a long measuring tape rather than a short one. A measured distance *D* can be viewed as the product of the length *l* of a measuring tape times the number *n* of times it was used. for instance, using a 3' tape 10 times gives a length of $30'$. To measure the same distance with a $12'$ tape, we would use the tape 2.5 times. (i.e., $30 = 12 \times 2.5$.) Thus $D = nl$.

Suppose each time a measurement is taken with the tape of the recorded distance within $1/16''$ of the actual distance, (i.e., $dl = 1/16'' \approx 0.005$ft). Using differentials, show why common sense proves correct in that it is better to use a long tape to measure long distances.

In Exercises 17-18, find the total differential dw.

17. $w = x^2yz^3$

18. $w = e^x \sin y \ln z$

In Exercises 19-22, use the information provided and the total differential to make the given approximation.

19. $f(3,1) = 7,\ f_x(3,1) = 9,\ f_y(3,1) = -2.$ Approximate $f(3.05, 0.9)$.

20. $f(-4,2) = 13,\ f_x(-4,2) = 2.6,\ f_y(-4,2) = 5.1.$ Approximate $f(-4.12, 2.07)$.

21. $f(2,4,5) = -1,\ f_x(2,4,5) = 2,\ f_y(2,4,5) = -3,\ f_z(2,4,5) = 3.7$ Approximate $f(2.5, 4.1, 4.8)$.

22. $f(3,3,3) = 5,\ f_x(3,3,3) = 2,\ f_y(3,3,3) = 0,\ f_z(3,3,3) = -2$ Approximate $f(3.1, 3.1, 3.1)$.

12.5: The Multivariable Chain Rule

Terms and Concepts

1. Let a level curve of $z = f(x,y)$ be described by $x = g(t),\ y = h(t)$. Explain why $\frac{dz}{dt} = 0$.

2. Fill in the blank: The single variable Chain Rule states $\frac{d}{dx}\left(f\left(g(x)\right)\right) = f'\left(g(x)\right) \cdot$___________.

3. Fill in the blank: The Multivariable Chain Rule states $\frac{df}{dt} = \frac{\partial f}{\partial x} \cdot$____________ + _____________$\cdot \frac{dy}{dt}$.

4. If $z = f(x,y)$, where $x = g(t)$ and $y = h(t)$, we can substitute and write *z* as an explicit function of *t*.
T/F: Using the Multivariable Chain Rule to find $\frac{dz}{dt}$ is sometimes easier than first substituting and then taking the derivative.

5. T/F: The Multivariable Chain Rule is only useful when all the related functions are known explicitly.

6. The Multivariable Chain Rule allows us to compute implicit derivatives easily by just computing two _______ derivatives.

Problems

In Exercises 7-12, functions $z = f(x,y)$, $x = g(t)$ and $y = h(t)$ are given.
(a) Use the Multivariable Chain Rule to compute $\frac{dz}{dt}$.
(b) Evaluate $\frac{dz}{dt}$ at the indicated *t*-value.

7. $z = 3x + 4y, \quad x = t^2, \quad y = 2t; \quad t = 1$

8. $z = x^2 - y^2, \quad x = t, \quad y = t^2 - 1; \quad t = 1$

9. $z = 5x + 2y, \quad x = 2\cos t + 1, \quad y = \sin t - 3; \quad t = \pi/4$

10. $z = \frac{x}{y^2+1}, \quad x = \cos t, \quad y = \sin t; \quad t = \pi/2$

11. $z = x^2 + 2y^2, \quad x = \sin t, \quad y = 3\sin t; \quad t = \pi/4$

12. $z = \cos x \sin y, \quad x = \pi t, \quad y = 2\pi t + \pi/2; \quad t = 3$

In Exercises 13-18, functions $z = f(x,y)$, $x = g(t)$ and $y = h(t)$ are given. Find the values of *t* where $\frac{dz}{dt} = 0$. Note: these are the same surfaces/curves as found in Exercises 7-12.

13. $z = 3x + 4y, \quad x = t^2, \quad y = 2t$

14. $z = x^2 - y^2, \quad x = t, \quad y = t^2 - 1$

15. $z = 5x + 2y, \quad x = 2\cos t + 1, \quad y = \sin t - 3$

16. $z = \frac{x}{y^2+1}, \quad x = \cos t, \quad y = \sin t$

17. $z = x^2 + 2y^2, \quad x = \sin t, \quad y = 3\sin t$

18. $z = \cos x \sin y, \quad x = \pi t, \quad y = 2\pi t + \pi/2$

In Exercises 19-22, functions $z = f(x, y), x = g(s, t)$ and $y = h(s, t)$ are given.
(a) Use the Multivariable Chain Rule to compute $\frac{\partial z}{\partial s}$ and \(\frac{\partial z}{\partial t}\).
(b) Evaluate $\frac{\partial z}{\partial s}$ and $\frac{\partial z}{\partial t}$ at the indicated *s* and *t* values.

19. $z = x^2y, \quad x = s - t, \quad y = 2s + 4t; \quad s = 1, t = 0$

20. $z = \cos\left(\pi x + \frac{\pi}{2}y\right), \quad x = st^2, \quad y = s^2t; \quad s = 1, t = 1$

21. $z = x^2 + y^2, \quad x = s\cos t, \quad y = s\sin t; \quad s = 2, t = \pi/4$

22. $z = e^{-(x^2+y^2)}, \quad x = t, \quad y = st^2; \quad s = 1, t = 1$

In Exercises 23-26, find $\frac{dy}{dx}$ using Implicit Differentiation and Theorem 109.

23. $x^2 \tan y = 50$

24. $(3x^2 + 2y^3)^4 = 2$

25. $\frac{x^2+y}{x+y^2} = 17$

26. $\ln(x^2 + xy + y^2) = 1$

In Exercises 27-30, find $\frac{dz}{dt}$, or $\frac{\partial z}{\partial s}$ and $\frac{\partial z}{\partial t}$, using the supplied information.

27. $\frac{\partial z}{\partial x} = 2, \quad \frac{\partial z}{\partial y} = 1, \quad \frac{dx}{dt} = 4, \quad \frac{dy}{dt} = -5$

28. $\frac{\partial z}{\partial x} = 1, \quad \frac{\partial z}{\partial y} = -3, \quad \frac{dx}{dt} = 6, \quad \frac{dy}{dt} = 2$

29. $\frac{\partial z}{\partial x} = -4, \quad \frac{\partial z}{\partial y} = 9, \quad \frac{\partial x}{\partial s} = 5, \quad \frac{\partial x}{\partial t} = 7, \quad \frac{\partial y}{\partial s} = -2, \quad \frac{\partial y}{\partial t} = 6$

30. $\frac{\partial z}{\partial x} = 2, \quad \frac{\partial z}{\partial y} = 1, \quad \frac{\partial x}{\partial s} = -2, \quad \frac{\partial x}{\partial t} = 3, \quad \frac{\partial y}{\partial s} = 2, \quad \frac{\partial y}{\partial t} = -1$

12.6: Directional Derivatives

Terms and Concepts

1. What is the difference between a directional derivative and a partial derivative?

2. For what $\vec{u}$ is $D_{\vec{u}}f = f_x$?

3. For what $\vec{u}$ is $D_{\vec{u}}f = f_y$?

4. The gradient is _______ to level curves.

5. The gradient points in the direction of _______ increase.

6. It is generally more informative to view the directional derivative not as the result of a limit, but rather as the result of a ________ product.

Problems

In Exercises 7-12, a function $z = f(x, y)$ is given. Find ∇f.

7. $f(x, y) = -x^2y + xy^2 + xy$

8. $f(x, y) = \sin x \cos y$

9. $f(x, y) = \frac{1}{x^2+y^2+1}$

10. $f(x, y) = -4x + 3y$

11. $f(x, y) = x^2 + 2y^3 - xy - 7x$

12. $f(x, y) = x^2y^3 - 2x$

In Exercises 13-18, a function $z = f(x, y)$ and a point *P* are given. Find the directional derivative of *f* in the indicated directions. Note: these are the same functions as in Exercises 7-12.

13. $f(x,y) = -x^2y + xy^2 + xy,\ P = (2,1)$
(a) In the direction of $\vec{v} = \langle 3,4 \rangle$
(b) In the direction toward the point $Q = (-1,1)$

14. $f(x,y) = \sin x \cos y,\ P = (\frac{\pi}{4}, \frac{\pi}{3})$
(a) In the direction of $\vec{v} = \langle 1,1 \rangle$
(b) In the direction toward the point $Q = (0,0)$

15. $f(x,y) = \frac{1}{x^2+y^2+1},\ P = (1,1)$
(a) In the direction of $\vec{v} = \langle 1,-1 \rangle$
(b) In the direction toward the point $Q = (-2,-2)$

16. $f(x,y) = -4x + 3y,\ P = (5,2)$
(a) In the direction of $\vec{v} = \langle 3,1 \rangle$
(b) In the direction toward the point $Q = (2,7)$

17. $f(x,y) = x^2 + 2y^2 - xy - 7x,\ P = (4,1)$
(a) In the direction of $\vec{v} = \langle -2,5 \rangle$
(b) In the direction toward the point $Q = (4,0)$

18. $f(x,y) = x^2y^3 - 2x,\ P = (1,1)$
(a) In the direction of $\vec{v} = \langle 3,3 \rangle$
(b) In the direction toward the point $Q = (1,2)$

In Exercises 19-24, a function $z = f(x,y)$ and a point *P* are given.
(a) Find the direction of maximal increase of f at P.
(b) What is the maximal value of $D_{\vec{u}}f$ at P?
(c) Find the direction of minimal increase of f at P.
(d) Give a direction $\vec{u}$ such that $D_{\vec{u}}f = 0$ at P.
Note: these are the same functions and points as in Exercises 13 through 18.

19. $f(x,y) = -x^2y + xy^2 + xy,\ P = (2,1)$

20. $f(x,y) = \sin x \cos y,\ P = (\frac{\pi}{4}, \frac{\pi}{3})$

21. $f(x,y) = \frac{1}{x^2+y^2+1},\ P = (1,1)$

22. $f(x,y) = -4x + 3y,\ P = (5,4)$

23. $f(x,y) = x^2 + 2y^2 - xy - 7x,\ P = (4,1)$

24. $f(x,y) = x^2y^3 - 2x,\ P = (1,1)$

In Exercises 25-28, a function $w = F(x,y,z)$, a vector $\vec{v}$ and a point *P* are given.
(a) Find $\nabla F(x,y,z)$.
(b) Find $D_{\vec{u}}F$ at *P*.

25. $f(x,y) = 3x^2z^3 + 4xy - 3z^2,\ \vec{v} = \langle 1,1,1 \rangle,\ P = (3,2,1)$

26. $f(x,y) = \sin(x)\cos(y)e^z,\ \vec{v} = \langle 2,2,1 \rangle,\ P = (0,0,0)$

27. $f(x,y) = x^2y^2 - y^2z^2,\ \vec{v} = \langle -1,7,3 \rangle,\ P = (1,0,-1)$

28. $f(x,y) = \frac{2}{x^2+y^2+z^2},\ \vec{v} = \langle 1,1,-2 \rangle,\ P = (1,1,1)$

12.7: Tangent Lines, Normal Lines, and Tangent Planes

Terms and Concepts

1. Explain how the vector $\vec{v} = \langle 1,0,3 \rangle$ can be thought of as having a "slope" of 3.

2. Explain how the vector $\vec{v} = \langle 0.6,0.8,-2 \rangle$ can be thought of as having a "slope" of -2.

3. T/F: Let $z = f(x,y)$ be differentiable at P. If $\vec{n}$ is normal vector to the tangent plane of f at P, then $\vec{n}$ is orthogonal to f_x and f_y at P.

4. Explain in your own words why we do not refer to the tangent line to a surface at a point, but rather to directional tangent lines to a surface at a point.

Problems

In Exercises 5-8, a function $z=f(x,y)$, a vector $\vec{v}$ and a point *P* are given. Give the parametric equations of the following directional tangent lines to f at P:
(a) $l_x(t)$
(b) $l_y(t)$
(c) $l_{\vec{u}}$**, where** $\vec{u}$ **is the unit vector in the direction of** $\vec{v}$.

5. $f(x,y)=2x^y-4xy^2,\ \vec{v}=\langle 1,3\rangle,\ P=(2,3).$

6. $f(x,y)=3\cos x\sin y,\ \vec{v}=\langle 1,2\rangle,\ P=(\pi/3,\pi/6).$

7. $f(x,y)=3x-5y,\ \vec{v}=\langle 1,1\rangle,\ P=(4,2).$

8. $f(x,y)=x^2-2x-y^2+4y,\ \vec{v}=\langle 1,1\rangle,\ P=(1,2).$

In Exercises 9-12, a function $z=f(x,y)$ and a point *P* are given. Find the equation of the normal line to f at P. Note: these are the same functions as in Exercises 5-8.

9. $f(x,y)=2x^y-4xy^2,\ \vec{v}=\langle 1,3\rangle,\ P=(2,3).$

10. $f(x,y)=3\cos x\sin y,\ \vec{v}=\langle 1,2\rangle,\ P=(\pi/3,\pi/6).$

11. $f(x,y)=3x-5y,\ \vec{v}=\langle 1,1\rangle,\ P=(4,2).$

12. $f(x,y)=x^2-2x-y^2+4y,\ \vec{v}=\langle 1,1\rangle,\ P=(1,2).$

In Exercises 13-16, a function $z=f(x,y)$ and a point *P* are given. Find the two points that are 2 units from the surface f at P. Note: these are the same functions as in Exercises 5-8.

13. $f(x,y)=2x^y-4xy^2,\ \vec{v}=\langle 1,3\rangle,\ P=(2,3).$

14. $f(x,y)=3\cos x\sin y,\ \vec{v}=\langle 1,2\rangle,\ P=(\pi/3,\pi/6).$

15. $f(x,y)=3x-5y,\ \vec{v}=\langle 1,1\rangle,\ P=(4,2).$

16. $f(x,y)=x^2-2x-y^2+4y,\ \vec{v}=\langle 1,1\rangle,\ P=(1,2).$

In Exercises 17-20, a function $z=f(x,y)$ and a point P are given. Find the equation of the tangent plane to f at P. Note: these are the same functions as in Exercises 5-8.

17. $f(x,y)=2x^y-4xy^2,\ \vec{v}=\langle 1,3\rangle,\ P=(2,3).$

18. $f(x,y)=3\cos x\sin y,\ \vec{v}=\langle 1,2\rangle,\ P=(\pi/3,\pi/6).$

19. $f(x,y)=3x-5y,\ \vec{v}=\langle 1,1\rangle,\ P=(4,2).$

20. $f(x,y)=x^2-2x-y^2+4y,\ \vec{v}=\langle 1,1\rangle,\ P=(1,2).$

In Exercises 21-24, an implicitly defined function of x, y and z is given along with a point P that lies on the surface. Use the gradient ∇F to:
(a) find the equation of the normal line to the surface at *P*, and
(b) find the equation of the plane tangent to the surface at *P*.

21. $\frac{x^2}{8}+\frac{y^2}{4}+\frac{z^2}{16}=1,\ \text{at } P=(1,\sqrt{2},\sqrt{6})$

22. $z^2-\frac{x^2}{4}-\frac{y^2}{9}=0,\ \text{at } P=(4,-3,\sqrt{5})$

23. $xy^2-xz^2=0,\ \text{at } P=(2,1,-1)$

24. $\sin(xy)+\cos)yz)=0,\ \text{at } P=(2,\pi/12,4)$

12.8: Extreme Values

Terms and Concepts

1. T/F: Theorem 114 states that if f has a critical point at P, then f has a relative extrema at P.

2. T/F: A point P is a critical point of f if f_x and f_y are both 0 at P.

3. T/F: A point P is a critical point of f if f_x or f_y are undefined at P.

4. Explain what it means to "solve a constrained optimization" problem.

Problems

Exercises 5-14, find the critical points of the given function. Use the Second Derivative Test to determine if each critical point corresponds to a relative maximum, minimum, or saddle point.

5. $f(x,y)=\frac{1}{2}x^2+2y^2-8y+4x$

6. $f(x,y)=x^2+4x+y^2-9y+3xy$

7. $f(x,y)=x^2+3y^2-6y+4xy$

8. $f(x,y)=\frac{1}{x^2+y^2+1}$

9. $f(x,y)=x^2+y^3-3y+1$

10. $f(x,y)=\frac{1}{3}x^3-x+\frac{1}{3}y^3-4y$

11. $f(x,y)=x^2y^2$

12. $f(x,y)=x^4-2x^2+y^3-27y-15$

13. $f(x,y)=\sqrt{16-(x-3)^2-y^2}$

14. $f(x,y)=\sqrt{x^2+y^2}$

In Exercises 15-18, find the absolute maximum and minimum of the function subject to the given constraint.

15. $f(x,y)=x^2+y^2+y=1$, constrained to the triangles with vertices $0,1),(-1,1),\text{ and }(1,-1)$.

16. $f(x,y)=5x-7y$, constrained to the region bounded by $y=x^2\text{ and }y=1$.

17. $f(x,y)=x^2+2x+y^2+2y$, constrained to the region bounded by the circle $x^2+y^2=4$.

18. $f(x,y)=3y-2x^2$, constrained to the region bounded by the parabola $y=x^2+x-1$ and the line $y=x$.

12.E: Applications of Functions of Several Variables (Exercises) is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

13: Multiple Integration

In this chapter we apply techniques of integral calculus to multivariable functions. In Chapter 5 we learned how the definite integral of a single variable function gave us "area under the curve." In this chapter we will see that integration applied to a multivariable function gives us "volume under a surface." And just as we learned applications of integration beyond finding areas, we will find applications of integration in this chapter beyond finding volume.

13.1: Iterated Integrals and Area
13.2: Double Integration and Volume
13.3: Double Integration with Polar Coordinates
13.4: Center of Mass
13.5: Surface Area
13.6: Volume Between Surfaces and Triple Integration
13.E: Applications of Multiple Integration (Exercises)

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 13: Multiple Integration is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

13.1: Iterated Integrals and Area

In the previous chapter we found that we could differentiate functions of several variables with respect to one variable, while treating all the other variables as constants or coefficients. We can integrate functions of several variables in a similar way. For instance, if we are told that $f_x(x,y)=2xy$, we can treat y as staying constant and integrate to obtain $f(x,y)$:

$$\begin{aligned} f(x,y) &= \int f_x(x,y)\,dx \\ &= \int 2xy\,dx \\ &= x^2y+C. \end{aligned}$$

Make a careful note about the constant of integration, C. This "constant" is something with a derivative of 0 with respect to x, so it could be any expression that contains only constants and functions of y. For instance, if $f(x,y)=x^2y+\sin y+y^3+17$, then $f_x(x,y)=2xy$. To signify that C is actually a function of y, we write:

$$f(x,y)=\int f_x(x,y)\,dx=x^2y+C(y). \tag{13.1.1}$$

Using this process we can even evaluate definite integrals.

Example 13.1.1: Integrating functions of more than one variable

Evaluate the integral $\displaystyle\int_1^{2y} 2xy\,dx$.

Solution

We find the indefinite integral as before, then apply the Fundamental Theorem of Calculus to evaluate the definite integral:

$$\begin{aligned} \int_1^{2y} 2xy\,dx &= x^2y\Big|_1^{2y} \\ &= (2y)^2y-(1)^2y \\ &= 4y^3-y. \end{aligned}$$

We can also integrate with respect to y. In general,

$$\int_{h_1(y)}^{h_2(y)} f_x(x,y)\,dx = f(x,y)\Big|_{h_1(y)}^{h_2(y)} = f\big(h_2(y),y\big)-f\big(h_1(y),y\big), \tag{13.1.2}$$

and

$$\int_{g_1(x)}^{g_2(x)} f_y(x,y)\,dy = f(x,y)\Big|_{g_1(x)}^{g_2(x)} = f\big(x,g_2(x)\big)-f\big(x,g_1(x)\big). \tag{13.1.3}$$

Note that when integrating with respect to x, the bounds are functions of y (of the form $x=h_1(y)$ and $x=h_2(y)$) and the final result is also a function of y. When integrating with respect to y, the bounds are functions of x (of the form $y=g_1(x)$ and $y=g_2(x)$) and the final result is a function of x. Another example will help us understand this.

Example 13.1.2: Integrating functions of more than one variable

Evaluate $\displaystyle\int_1^{x}\left(5x^3y^{-3}+6y^2\right)dy$.

Solution

We consider x as staying constant and integrate with respect to y:

$$\begin{aligned}\int_1^x \left(5x^3y^{-3}+6y^2\right)dy &= \left(\frac{5x^3y^{-2}}{-2}+\frac{6y^3}{3}\right)\Bigg|_1^x \\ &= \left(-\frac{5}{2}x^3x^{-2}+2x^3\right)-\left(-\frac{5}{2}x^3+2\right) \\ &= \frac{9}{2}x^3-\frac{5}{2}x-2.\end{aligned}$$

Note how the bounds of the integral are from $y=1$ to $y=x$ and that the final answer is a function of x.

In the previous example, we integrated a function with respect to y and ended up with a function of x. We can integrate this as well. This process is known as **iterated integration**, or **multiple integration**.

Example 13.1.3: Integrating an integral

Evaluate $\displaystyle\int_1^2\left(\int_1^x\left(5x^3y^{-3}+6y^2\right)dy\right)dx$.

Solution

We follow a standard "order of operations" and perform the operations inside parentheses first (which is the integral evaluated in Example 13.1.2.)

$$\begin{aligned}\int_1^2\left(\int_1^x\left(5x^3y^{-3}+6y^2\right)dy\right)dx &= \int_1^2\left(\left[\frac{5x^3y^{-2}}{-2}+\frac{6y^3}{3}\right]\Bigg|_1^x\right)dx \\ &= \int_1^2\left(\frac{9}{2}x^3-\frac{5}{2}x-2\right)dx \\ &= \left(\frac{9}{8}x^4-\frac{5}{4}x^2-2x\right)\Bigg|_1^2 \\ &= \frac{89}{8}.\end{aligned}$$

Note how the bounds of x were $x=1$ to $x=2$ and the final result was a number.

The previous example showed how we could perform something called an iterated integral; we do not yet know why we would be interested in doing so nor what the result, such as the number $89/8$, *means*. Before we investigate these questions, we offer some definitions.

Definition: Iterated Integration

Iterated integration is the process of repeatedly integrating the results of previous integrations. Integrating one integral is denoted as follows.

Let a, b, c and d be numbers and let $g_1(x)$, $g_2(x)$, $h_1(y)$ and $h_2(y)$ be functions of x and y, respectively. Then:

1. $\displaystyle\int_c^d\int_{h_1(y)}^{h_2(y)} f(x,y)\,dx\,dy=\int_c^d\left(\int_{h_1(y)}^{h_2(y)} f(x,y)\,dx\right)dy.$
2. $\displaystyle\int_a^b\int_{g_1(x)}^{g_2(x)} f(x,y)\,dy\,dx=\int_a^b\left(\int_{g_1(x)}^{g_2(x)} f(x,y)\,dy\right)dx.$

Again make note of the bounds of these iterated integrals.

With $\displaystyle\int_c^d\int_{h_1(y)}^{h_2(y)} f(x,y)\,dx\,dy$, x varies from $h_1(y)$ to $h_2(y)$, whereas y varies from c to d. That is, the bounds of x are *curves*, the curves $x=h_1(y)$ and $x=h_2(y)$, whereas the bounds of y are *constants*, $y=c$ and $y=d$. It is useful to remember that when setting up and evaluating such iterated integrals, we integrate "from curve to curve, then from point to point."

We now begin to investigate *why* we are interested in iterated integrals and *what* they mean.

Area of a plane region

Consider the plane region R bounded by $a\le x\le b$ and $g_1(x)\le y\le g_2(x)$, shown in Figure 13.1.1. We learned in Section 7.1 (in Calculus I) that the area of R is given by

$$\int_a^b \big(g_2(x)-g_1(x)\big)\,dx. \tag{13.1.4}$$

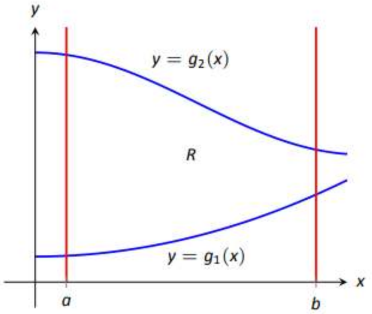


Figure 13.1.1: Calculating the area of a plane region R with an integrated integral.

We can view the expression $\big(g_2(x)-g_1(x)\big)$ as

$$\big(g_2(x)-g_1(x)\big)=\int_{g_1(x)}^{g_2(x)} 1\,dy=\int_{g_1(x)}^{g_2(x)}\,dy,$$

meaning we can express the area of R as an iterated integral:

$$\text{area of } R=\int_a^b \big(g_2(x)-g_1(x)\big)\,dx=\int_a^b\left(\int_{g_1(x)}^{g_2(x)}\,dy\right)\,dx=\int_a^b\int_{g_1(x)}^{g_2(x)}\,dy\,dx. \tag{13.1.5}$$

In short: a certain iterated integral can be viewed as giving the area of a plane region.

A region R could also be defined by $c\le y\le d$ and $h_1(y)\le x\le h_2(y)$, as shown in Figure 13.1.2. Using a process similar to that above, we have

$$\text{the area of } R=\int_c^d\int_{h_1(y)}^{h_2(y)}\,dx\,dy. \tag{13.1.6}$$

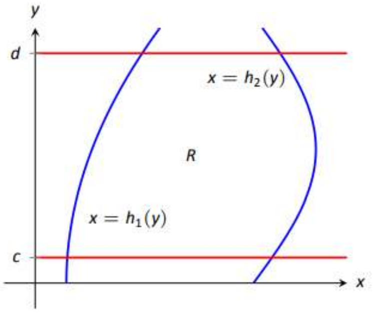


Figure 13.1.2: Calculating the area of a plane region R with an iterated integral.

We state this formally in a theorem.

THEOREM 13.1.1: Area of a plane region

1. Let R be a plane region bounded by $a\le x\le b$ and $g_1(x)\le y\le g_2(x)$, where g_1 and g_2 are continuous functions on $[a,b]$. The area A of R is

$$A=\int_a^b\int_{g_1(x)}^{g_2(x)}\,dy\,dx. \tag{13.1.7}$$

2. Let R be a plane region bounded by $c\le y\le d$ and $h_1(y)\le x\le h_2(y)$, where h_1 and h_2 are continuous functions on $[c,d]$. The area A of R is

$$A=\int_c^d\int_{h_1(y)}^{h_2(y)}\,dx\,dy. \tag{13.1.8}$$

The following examples should help us understand this theorem.

Example 13.1.4: Area of a rectangle

Find the area A of the rectangle with corners $(-1,1)$ and $(3,3)$, as shown in Figure 13.1.3.

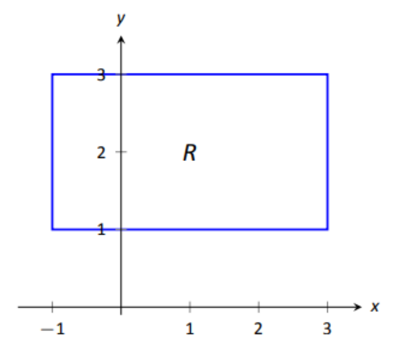


Figure 13.1.3: Calculating the area of a rectangle with an iterated integral in Example 13.1.4.

Solution

Multiple integration is obviously overkill in this situation, but we proceed to establish its use.

The region R is bounded by $x=-1$, $x=3$, $y=1$ and $y=3$. Choosing to integrate with respect to y first, we have

$$A=\int_{-1}^{3}\int_{1}^{3}1\,dy\,dx=\int_{-1}^{3}\left(y\Big|_{1}^{3}\right)\,dx=\int_{-1}^{3}2\,dx=2x\Big|_{-1}^{3}=8.$$

We could also integrate with respect to x first, giving:

$$A=\int_{1}^{3}\int_{-1}^{3}1\,dx\,dy=\int_{1}^{3}\left(x\Big|_{-1}^{3}\right)\,dy=\int_{1}^{3}4\,dy=4y\Big|_{1}^{3}=8.$$

Clearly there are simpler ways to find this area, but it is interesting to note that this method works.

Example 13.1.5: Area of a triangle

Find the area A of the triangle with vertices at $(1,1)$, $(3,1)$ and $(5,5)$, as shown in Figure 13.1.4.

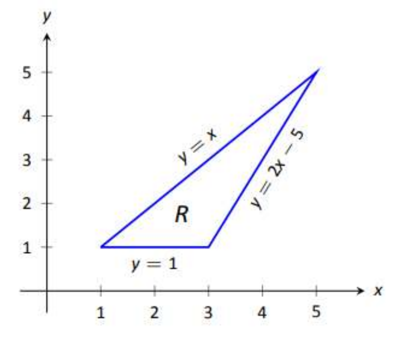


Figure 13.1.4: Calculating the area of a triangle with iterated integrals in Example 13.1.5.

Solution

The triangle is bounded by the lines as shown in the figure. Choosing to integrate with respect to x first gives that x is bounded by $x=y$ to $x=\frac{y+5}{2}$, while y is bounded by $y=1$ to $y=5$. (Recall that since x-values increase from left to right, the leftmost curve, $x=y$, is the lower bound and the rightmost curve, $x=(y+5)/2$, is the upper bound.) The area is

$$\begin{aligned}A &= \int_1^5 \int_y^{\frac{y+5}{2}} dx\,dy \\ &= \int_1^5 \left(x\,\Big|_y^{\frac{y+5}{2}}\right) dy \\ &= \int_1^5 \left(-\frac{1}{2}y+\frac{5}{2}\right) dy \\ &= \left(-\frac{1}{4}y^2+\frac{5}{2}y\right)\Big|_1^5 \\ &= 4.\end{aligned}$$

We can also find the area by integrating with respect to y first. In this situation, though, we have two functions that act as the lower bound for the region R, $y=1$ and $y=2x-5$. This requires us to use two iterated integrals. Note how the x-bounds are different for each integral:

$$\begin{aligned}A &= \int_1^3 \int_1^x 1\,dy\,dx && + && \int_3^5 \int_{2x-5}^x 1\,dy\,dx \\ &= \int_1^3 (y)\Big|_1^x dx && + && \int_3^5 (y)\Big|_{2x-5}^x dx \\ &= \int_1^3 (x-1)\,dx && + && \int_3^5 (-x+5)\,dx \\ &= 2 && + && 2 \\ &= 4.\end{aligned}$$

As expected, we get the same answer both ways.

Example 13.1.6: Area of a plane region

Find the area of the region enclosed by $y=2x$ and $y=x^2$, as shown in Figure 13.1.5.

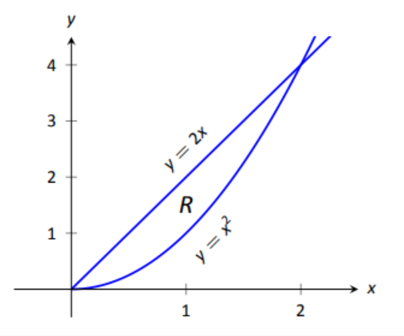


Figure 13.1.5: Calculating the area of a plane region with iterated integrals in Example 13.1.6.

Solution

Once again we'll find the area of the region using both orders of integration.

Using $dy\,dx$:

$$\int_0^2 \int_{x^2}^{2x} 1\,dy\,dx = \int_0^2 (2x-x^2)\,dx = \left(x^2-\frac{1}{3}x^3\right)\Big|_0^2 = \frac{4}{3}. \tag{13.1.9}$$

Using $dx\,dy$:

$$\int_0^4 \int_{y/2}^{\sqrt{y}} 1\,dx\,dy = \int_0^4 (\sqrt{y}-y/2)\,dy = \left(\frac{2}{3}y^{3/2}-\frac{1}{4}y^2\right)\Big|_0^4 = \frac{4}{3}. \tag{13.1.10}$$

Changing Order of Integration

In each of the previous examples, we have been given a region R and found the bounds needed to find the area of R using both orders of integration. We integrated using both orders of integration to demonstrate their equality.

We now approach the skill of describing a region using both orders of integration from a different perspective. Instead of starting with a region and creating iterated integrals, we will start with an iterated integral and rewrite it in the other integration order. To do so, we'll need to understand the region over which we are integrating.

The simplest of all cases is when both integrals are bound by constants. The region described by these bounds is a rectangle (see Example $13.1.4$), and so:

$$\int_a^b \int_c^d 1\,dy\,dx = \int_c^d \int_a^b 1\,dx\,dy. \tag{13.1.11}$$

When the inner integral's bounds are not constants, it is generally very useful to sketch the bounds to determine what the region we are integrating over looks like. From the sketch we can then rewrite the integral with the other order of integration.

Examples will help us develop this skill.

Example $13.1.7$: Changing the order of integration

Rewrite the iterated integral $\displaystyle\int_0^6 \int_0^{x/3} 1\,dy\,dx$ with the order of integration $dx\,dy$.

Solution

We need to use the bounds of integration to determine the region we are integrating over.

The bounds tell us that y is bounded by 0 and $x/3$; x is bounded by 0 and 6. We plot these four curves: $y=0$, $y=x/3$, $x=0$ and $x=6$ to find the region described by the bounds. Figure $13.1.6$ shows these curves, indicating that R is a triangle.

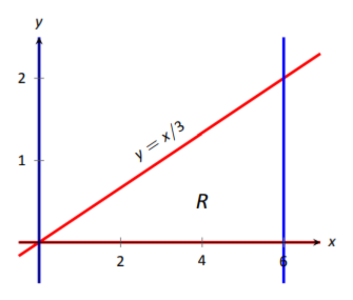


Figure $13.1.6$: Sketching the region R described by the iterated integral in Example $13.1.7$.

To change the order of integration, we need to consider the curves that bound the x-values. We see that the lower bound is $x = 3y$ and the upper bound is $x=6$. The bounds on y are 0 to 2. Thus we can rewrite the integral as $\displaystyle\int_0^2 \int_{3y}^6 1\,dx\,dy.$

Example $13.1.8$: Changing the order of integration

Change the order of integration of $\displaystyle\int_0^4 \int_{y^2/4}^{(y+4)/2} 1\,dx\,dy.$

Solution

We sketch the region described by the bounds to help us change the integration order. x is bounded below and above (i.e., to the left and right) by $x = y^2/4$ and $x = (y+4)/2$ respectively, and y is bounded between 0 and 4. Graphing the previous curves, we find the region R to be that shown in Figure $13.1.7$.

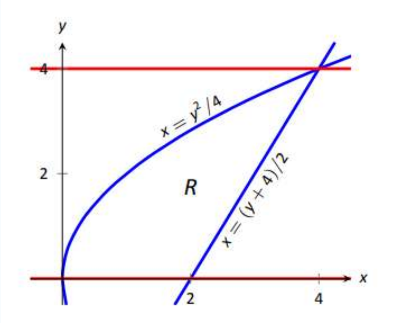


Figure 13.1.7: Drawing the region determined by the bounds of integration in Example 13.1.8.

To change the order of integration, we need to establish curves that bound y. The figure makes it clear that there are two lower bounds for y: $y=0$ on $0 \le x \le 2$, and $y = 2x-4$ on $2 \le x \le 4$. Thus we need two double integrals. The upper bound for each is $y = 2\sqrt{x}$. Thus we have

$$\int_0^4 \int_{y^2/4}^{(y+4)/2} 1\,dx\,dy = \int_0^2 \int_0^{2\sqrt{x}} 1\,dy\,dx + \int_2^4 \int_{2x-4}^{2\sqrt{x}} 1\,dy\,dx.$$

This section has introduced a new concept, the iterated integral. We developed one application for iterated integration: area between curves. However, this is not new, for we already know how to find areas bounded by curves.

In the next section we apply iterated integration to solve problems we currently do not know how to handle. The "real" goal of this section was not to learn a new way of computing area. Rather, our goal was to learn how to define a region in the plane using the bounds of an iterated integral. That skill is very important in the following sections.

This page titled 13.1: Iterated Integrals and Area is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

13.2: Double Integration and Volume

The definite integral of f over $[a,b]$, $\int_a^b f(x)\,dx$, was introduced as "the signed area under the curve." We approximated the value of this area by first subdividing $[a,b]$ into n subintervals, where the $i^{\text{ th}}$ subinterval has length Δx_i, and letting c_i be any value in the $i^{\text{ th}}$ subinterval. We formed rectangles that approximated part of the region under the curve with width Δx_i, height $f(c_i)$, and hence with area $f(c_i)\,\Delta x_i$. Summing all the rectangle's areas gave an approximation of the definite integral, and Theorem 38 stated that

$$\int_a^b f(x)\,dx = \lim_{\|\Delta x\|\to 0} \sum f(c_i)\,\Delta x_i, \tag{13.2.1}$$

connecting the area under the curve with sums of the areas of rectangles.

We use a similar approach in this section to find volume under a surface.

Let R be a closed, bounded region in the xy-plane and let $z=f(x,y)$ be a continuous function defined on R. We wish to find the signed volume under the surface of f over R. (We use the term "signed volume" to denote that space above the xy-plane, under f, will have a positive volume; space above f and under the xy-plane will have a "negative" volume, similar to the notion of signed area used before.)

We start by partitioning R into n rectangular subregions as shown in Figure $13.2.1a$. For simplicity's sake, we let all widths be Δx and all heights be Δy. Note that the sum of the areas of the rectangles is not equal to the area of R, but rather is a close approximation. Arbitrarily number the rectangles 1 through n, and pick a point $(x_i,\, y_i)$ in the $i^{\text{ th}}$ subregion.

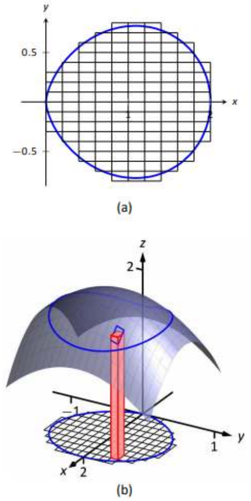


Figure $13.2.1$: Developing a method for finding signed volume under a surface.

The volume of the rectangular solid whose base is the i^{th} subregion and whose height is $f(x_i,\, y_i)$ is $V_i = f(x_i,\, y_i)\,\Delta x\,\Delta y$. Such a solid is shown in Figure $13.2.1b$. Note how this rectangular solid only approximates the true volume under the surface; part of the solid is above the surface and part is below.

For each subregion R_i used to approximate R, create the rectangular solid with base area $\Delta x\,\Delta y$ and height $f(x_i,\, y_i)$. The sum of all rectangular solids is

$$\sum_{i=1}^{n} f(x_i, y_i)\,\Delta x\,\Delta y. \tag{13.2.2}$$

This approximates the signed volume under f over R. As we have done before, to get a better approximation we can use more rectangles to approximate the region R.

In general, each rectangle could have a different width Δx_j and height Δy_k, giving the $i^{\text{ th}}$ rectangle an area $\Delta A_i = \Delta x_j\,\Delta y_k$ and the i^{th} rectangular solid a volume of $f(x_i, y_i)\,\Delta A_i$. Let $||\Delta A||$ denote the length of the longest diagonal of all rectangles in the subdivision of R; $||\Delta A|| \to 0$ means each rectangle's width and height are both approaching 0. If f is a continuous function, as

$||\Delta A||$ shrinks (and hence $n \to \infty$) the summation $\sum_{i=1}^{n} f(x_i, y_i)\,\Delta A_i$ approximates the signed volume better and better. This leads to a definition.

Note: Recall that the integration symbol "$\int$" is an "elongated S," representing the word "sum." We interpreted $\int_a^b f(x)\,dx$ as "take the *sum* of the areas of rectangles over the interval $[a, b]$." The double integral uses two integration symbols to represent a "double sum." When adding up the volumes of rectangular solids over a partition of a region R, as done in Figure 13.2.1, one could first add up the volumes across each row (one type of sum), then add these totals together (another sum), as in

$$\sum_{j=1}^{n} \sum_{i=1}^{m} f(x_i, y_j)\,\Delta x_i\,\Delta y_j. \tag{13.2.3}$$

One can rewrite this as

$$\sum_{j=1}^{n} \left(\sum_{i=1}^{m} f(x_i, y_j)\,\Delta x_i \right) \Delta y_j. \tag{13.2.4}$$

The summation inside the parenthesis indicates the sum of heights × widths, which gives an area; multiplying these areas by the thickness Δy_j gives a volume. The illustration in Figure 13.2.2 relates to this understanding.

Definition 101: Double Integral, Signed Volume

Let $z = f(x, y)$ be a continuous function defined over a closed region R in the xy-plane. The **signed volume** V under f over R is denoted by the **double integral**

$$V = \iint_R f(x, y)\,dA. \tag{13.2.5}$$

Alternate notations for the double integral are

$$\iint_R f(x, y)\,dA = \iint_R f(x, y)\,dx\,dy = \iint_R f(x, y)\,dy\,dx. \tag{13.2.6}$$

The definition above does not state how to find the signed volume, though the notation offers a hint. We need the next two theorems to evaluate double integrals to find volume.

theorem 118: Double Integrals and Signed Volume

Let $z = f(x, y)$ be a continuous function defined over a closed region R in the xy-plane. Then the signed volume V under f over R is

$$V = \iint_R f(x, y) dA = \lim_{||\Delta A|| \to 0} \sum_{i=1}^{n} f(x_i, y_i)\,\Delta A_i. \tag{13.2.7}$$

This theorem states that we can find the exact signed volume using a limit of sums. The partition of the region R is not specified, so any partitioning where the diagonal of each rectangle shrinks to 0 results in the same answer.

This does not offer a very satisfying way of computing area, though. Our experience has shown that evaluating the limits of sums can be tedious. We seek a more direct method.

Recall Theorem 54 in Section 7.2. This stated that if $A(x)$ gives the cross-sectional area of a solid at x, then $\int_a^b A(x) dx$ gave the volume of that solid over $[a, b]$.

Consider Figure 13.2.2, where a surface $z = f(x, y)$ is drawn over a region R. Fixing a particular x value, we can consider the area under f over R where x has that fixed value. That area can be found with a definite integral, namely

$$A(x) = \int_{g_1(x)}^{g_2(x)} f(x, y)\,dy. \tag{13.2.8}$$

Remember that though the integrand contains x, we are viewing x as fixed. Also note that the bounds of integration are functions of x: the bounds depend on the value of x.

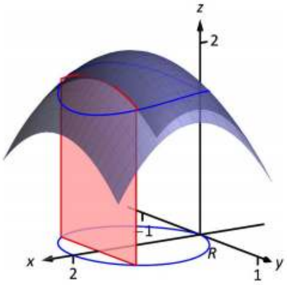


Figure 13.2.2: Finding volume under a surface by sweeping out a cross-sectional area.

As $A(x)$ is a cross-sectional area function, we can find the signed volume V under f by integrating it:

$$V = \int_a^b A(x)\,dx = \int_a^b \left(\int_{g_1(x)}^{g_2(x)} f(x,y)\,dy \right)\,dx = \int_a^b \int_{g_1(x)}^{g_2(x)} f(x,y)\,dy\,dx. \tag{13.2.9}$$

This gives a concrete method for finding signed volume under a surface. We could do a similar procedure where we started with y fixed, resulting in a iterated integral with the order of integration $dx\,dy$. The following theorem states that both methods give the same result, which is the value of the double integral. It is such an important theorem it has a name associated with it.

THEOREM 119: Fubini's Theorem

Let R be a closed, bounded region in the xy-plane and let $z = f(x,y)$ be a continuous function on R.

1. If R is bounded by $a \le x \le b$ and $g_1(x) \le y \le g_2(x)$, where g_1 and g_2 are continuous functions on $[a,b]$, then

$$\iint_R f(x,y)\,dA = \int_a^b \int_{g_1(x)}^{g_2(x)} f(x,y)\,dy\,dx. \tag{13.2.10}$$

2. If R is bounded by $c \le y \le d$ and $h_1(y) \le x \le h_2(y)$, where h_1 and h_2 are continuous functions on $[c,d]$, then

$$\iint_R f(x,y)\,dA = \int_c^d \int_{h_1(y)}^{h_2(y)} f(x,y)\,dx\,dy. \tag{13.2.11}$$

Note that once again the bounds of integration follow the "curve to curve, point to point" pattern discussed in the previous section. In fact, one of the main points of the previous section is developing the skill of describing a region R with the bounds of an iterated integral. Once this skill is developed, we can use double integrals to compute many quantities, not just signed volume under a surface.

Example 13.2.1: Evaluating a double integral

Let $f(x,y) = xy + e^y$. Find the signed volume under f on the region R, which is the rectangle with corners $(3,1)$ and $(4,2)$ pictured in Figure 13.2.3, using Fubini's Theorem and both orders of integration.

Solution

We wish to evaluate $\iint_R \left(xy + e^y\right)\,dA$. As R is a rectangle, the bounds are easily described as $3 \le x \le 4$ and $1 \le y \le 2$.

Using the order $dy\,dx$:

$$\begin{aligned}\iint_R \left(xy + e^y\right)\,dA &= \int_3^4 \int_1^2 \left(xy + e^y\right)\,dy\,dx \\ &= \int_3^4 \left(\left[\frac{1}{2}xy^2 + e^y\right]\Bigg|_1^2 \right)\,dx \\ &= \int_3^4 \left(\frac{3}{2}x + e^2 - e\right)\,dx \\ &= \left(\frac{3}{4}x^2 + \left(e^2 - e\right)x\right)\Bigg|_3^4 \\ &= \frac{21}{4} + e^2 - e \approx 9.92.\end{aligned}$$

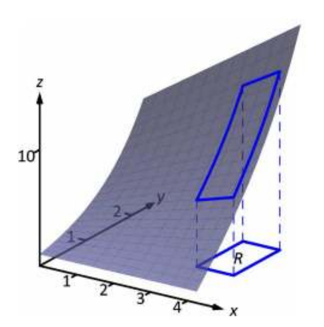


Figure 13.2.3: Finding the signed volume under a surface in Example 13.2.1.

Now we check the validity of Fubini's Theorem by using the order $dx\,dy$:

$$\begin{aligned}\iint_R \left(xy+e^y\right) dA &= \int_1^2 \int_3^4 \left(xy+e^y\right) dx\,dy \\ &= \int_1^2 \left(\left[\frac{1}{2}x^2y+xe^y\right]\Big|_3^4\right) dy \\ &= \int_1^2 \left(\frac{7}{2}y+e^y\right) dy \\ &= \left(\frac{7}{4}y^2+e^y\right)\Big|_1^2 \\ &= \frac{21}{4}+e^2-e \approx 9.92.\end{aligned}$$

Both orders of integration return the same result, as expected.

Example 13.2.2: Evaluating a double integral

Evaluate $\iint_R \left(3xy-x^2-y^2+6\right) dA$, where R is the triangle bounded by $x=0$, $y=0$ and $x/2+y=1$, as shown in Figure 13.2.4.

Solution

While it is not specified which order we are to use, we will evaluate the double integral using both orders to help drive home the point that it does not matter which order we use.

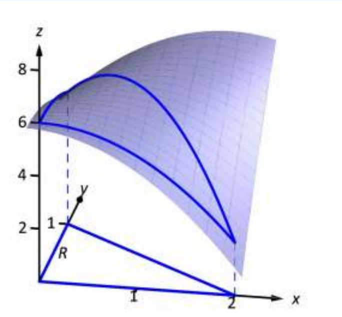


Figure 13.2.4: Finding the signed volume under the surface in Example 13.2.2.

Using the order $dy\,dx$:

The bounds on y go from "curve to curve," i.e., $0 \le y \le 1-x/2$, and the bounds on x go from "point to point," i.e., $0 \le x \le 2$.

$$\begin{aligned}\iint_R \left(3xy - x^2 - y^2 + 6\right) dA &= \int_0^2 \int_0^{-\frac{x}{2}+1} \left(3xy - x^2 - y^2 + 6\right) dy\, dx \\ &= \int_0^2 \left(\frac{3}{2}xy^2 - x^2y - \frac{1}{3}y^3 + 6y\right)\Bigg|_0^{-\frac{x}{2}+1} dx \\ &= \int_0^2 \left(\frac{11}{12}x^3 - \frac{11}{4}x^2 - x - \frac{17}{3}\right) dx \\ &= \left(\frac{11}{48}x^4 - \frac{11}{12}x^3 - \frac{1}{2}x^2 - \frac{17}{3}x\right)\Bigg|_0^2 \\ &= \frac{17}{3} = 5.\overline{6}.\end{aligned}$$

Now lets consider the order $dx dy$. Here x goes from "curve to curve," $0 \leq x \leq 2 - 2y$, and y goes from "point to point," $0 \leq y \leq 1$:

$$\begin{aligned}\iint_R \left(3xy - x^2 - y^2 + 6\right) dA &= \int_0^1 \int_0^{2-2y} \left(3xy - x^2 - y^2 + 6\right) dx\, dy \\ &= \int_0^1 \left(\frac{3}{2}x^2y - \frac{1}{3}x^3 - xy^2 + 6x\right)\Bigg|_0^{2-2y} dy \\ &= \int_0^1 \left(\frac{32}{3}y^3 - 22y^2 + 2y + \frac{28}{3}\right) dy \\ &= \left(\frac{8}{3}y^4 - \frac{22}{3}y^3 + y^2 + \frac{28}{3}y\right)\Bigg|_0^1 \\ &= \frac{17}{3} = 5.\overline{6}.\end{aligned}$$

We obtained the same result using both orders of integration.

Note how in these two examples that the bounds of integration depend only on R; the bounds of integration have nothing to do with $f(x, y)$. This is an important concept, so we include it as a Key Idea.

KEY IDEA 56: Double Integration Bounds

When evaluating $\iint_R f(x, y) dA$ using an iterated integral, the bounds of integration depend only on R. The surface f does not determine the bounds of integration.

Before doing another example, we give some properties of double integrals. Each should make sense if we view them in the context of finding signed volume under a surface, over a region.

THEOREM 120 Properties of Double Integrals

Let f and g be continuous functions over a closed, bounded plane region R, and let c be a constant.

1. $\iint_R c\, f(x, y) dA = c \iint_R f(x, y)\, dA.$
2. $\iint_R \left(f(x, y) \pm g(x, y)\right) dA = \iint_R f(x, y)\, dA \pm \iint_R g(x, y)\, dA$
3. If $f(x, y) \geq 0$ on R, then $\iint_R f(x, y)\, dA \geq 0.$
4. If $f(x, y) \geq g(x, y)$ on R, then $\iint_R f(x, y)\, dA \geq \iint_R g(x, y)\, dA.$
5. Let R be the union of two nonoverlapping regions, $R = R_1 \bigcup R_2$ (see Figure $13.2.5$). Then

$$\iint_R f(x, y)\, dA = \iint_{R_1} f(x, y)\, dA + \iint_{R_2} f(x, y)\, dA. \tag{13.2.12}$$

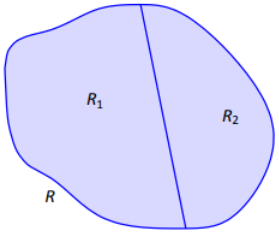


Figure 13.2.5: R is the union of two nonoverlapping regions, R_1 and R_2.

Example 13.2.3: Evaluating a double integral

Let $f(x, y) = \sin x \cos y$ and R be the triangle with vertices $(-1, 0)$, $(1, 0)$ and $(0, 1)$ (Figure 13.2.5). Evaluate the double integral $\iint_R f(x, y)\, dA$.

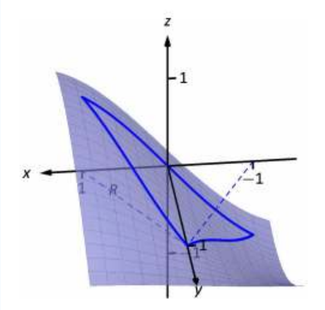


Figure 13.2.5: Finding the signed volume under a surface in Example 13.2.3.

Solution

If we attempt to integrate using an iterated integral with the order $dydx$, note how there are two upper bounds on R meaning we'll need to use two iterated integrals. We would need to split the triangle into two regions along the y-axis, then use Theorem 120, part 5.

Instead, let's use the order $dxdy$. The curves bounding x are $y-1 \leq x \leq 1-y$; the bounds on y are $0 \leq y \leq 1$. This gives us:

$$\begin{aligned}\iint_R f(x, y)\, dA &= \int_0^1 \int_{y-1}^{1-y} \sin x \cos y\, dx\, dy \\ &= \int_0^1 \Big(-\cos x \cos y \Big)\Big|_{y-1}^{1-y}\, dy \\ &= \int_0^1 \cos y \Big(-\cos(1-y) + \cos(y-1) \Big)\, dy.\end{aligned}$$

Recall that the cosine function is an even function; that is, $\cos x = \cos(-x)$. Therefore, from the last integral above, we have $\cos(y-1) = \cos(1-y)$. Thus the integrand simplifies to 0, and we have

$$\begin{aligned}\iint_R f(x, y)\, dA &= \int_0^1 0\, dy \\ &= 0.\end{aligned}$$

It turns out that over R, there is just as much volume above the xy-plane as below (look again at Figure 13.2.5), giving a final signed volume of 0.

Example 13.2.4: Evaluating a double integral

Evaluate $\iint_R (4-y)\, dA$, where R is the region bounded by the parabolas $y^2 = 4x$ and $x^2 = 4y$, graphed in Figure 13.2.6.

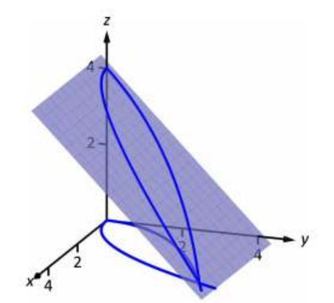


Figure 13.2.6: Finding the volume under the surface in Example 13.2.4.

Solution

Graphing each curve can help us find their points of intersection. Solving analytically, the second equation tells us that $y = x^2/4$. Substituting this value in for y in the first equation gives us $x^4/16 = 4x$. Solving for x:

$$\begin{aligned} \frac{x^4}{16} &= 4x \\ x^4 - 64x &= 0 \\ x(x^3-64) &= 0 \\ x &= 0,\ 4. \end{aligned}$$

Thus we've found analytically what was easy to approximate graphically: the regions intersect at $(0,0)$ and $(4,4)$, as shown in Figure 13.2.6.

We now choose an order of integration: $dy\,dx$ or $dx\,dy$? Either order works; since the integrand does not contain x, choosing $dx\,dy$ might be simpler -- at least, the first integral is very simple.

Thus we have the following "curve to curve, point to point" bounds: $y^2/4 \le x \le 2\sqrt{y}$, and $0 \le y \le 4$.

$$\begin{aligned} \iint_R (4-y)\,dA &= \int_0^4 \int_{y^2/4}^{2\sqrt{y}} (4-y)\,dx\,dy \\ &= \int_0^4 \Big(x(4-y)\Big)\Big|_{y^2/4}^{2\sqrt{y}}\,dy \\ &= \int_0^4 \Big(\big(2\sqrt{y}-\frac{y^2}{4}\big)(4-y)\Big)\,dy = \int_0^4 \Big(\frac{y^3}{4} - y^2 - 2y^{3/2} + 8y^{1/2}\Big)\,dy \\ &= \left(\frac{y^4}{16} - \frac{y^3}{3} - \frac{4y^{5/2}}{5} + \frac{16y^{3/2}}{3}\right)\Bigg|_0^4 \\ &= \frac{176}{15} = 11.7\overline{3}. \end{aligned}$$

The signed volume under the surface f is about 11.7 cubic units.

In the previous section we practiced changing the order of integration of a given iterated integral, where the region R was not explicitly given. Changing the bounds of an integral is more than just an test of understanding. Rather, there are cases where integrating in one order is really hard, if not impossible, whereas integrating with the other order is feasible.

Example 13.2.5: Changing the order of integration

Rewrite the iterated integral $\int_0^3 \int_y^3 e^{-x^2}\,dx\,dy$ with the order $dy\,dx$. Comment on the feasibility to evaluate each integral.

Solution

Once again we make a sketch of the region over which we are integrating to facilitate changing the order. The bounds on x are from $x = y$ to $x = 3$; the bounds on y are from $y = 0$ to $y = 3$. These curves are sketched in Figure 13.2.7, enclosing the region R.

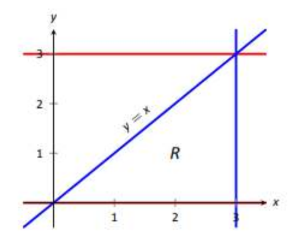


Figure 13.2.7: Determining the region R determined by the bounds of integration in Example 13.2.5.

To change the bounds, note that the curves bounding y are $y = 0$ up to $y = x$; the triangle is enclosed between $x = 0$ and $x = 3$. Thus the new bounds of integration are $0 \le y \le x$ and $0 \le x \le 3$, giving the iterated integral $\int_0^3 \int_0^x e^{-x^2}\,dy\,dx$.

How easy is it to evaluate each iterated integral? Consider the order of integrating $dxdy$, as given in the original problem. The first indefinite integral we need to evaluate is $\int e^{-x^2}\,dx$; we have stated before (see Section 5.5) that this integral cannot be evaluated in terms of elementary functions. We are stuck.

Changing the order of integration makes a big difference here. In the second iterated integral, we are faced with $\int e^{-x^2}\,dy$; integrating with respect to y gives us $ye^{-x^2} + C$, and the first definite integral evaluates to

$$\int_0^x e^{-x^2}\,dy = xe^{-x^2}.$$

Thus

$$\int_0^3 \int_0^x e^{-x^2}\,dy\,dx = \int_0^3 \left(xe^{-x^2}\right)\,dx.$$

This last integral is easy to evaluate with substitution, giving a final answer of $\frac{1}{2}(1 - e^{-9}) \approx 0.5$. Figure 13.2.8 shows the surface over R.

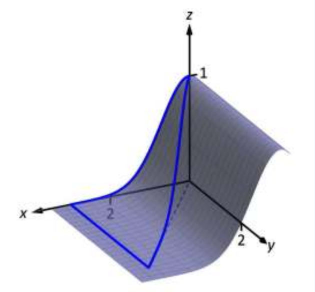


Figure 13.2.8: Showing the surface f defined in Example 13.2.5 over its region R.

In short, evaluating one iterated integral is impossible; the other iterated integral is relatively simple.

Definition 22 defines the average value of a single--variable function $f(x)$ on the interval $[a, b]$ as

$$\text{average value of } f(x) \text{ on } [a,b] = \frac{1}{b-a}\int_a^b f(x)\,dx;$$

that is, it is the "area under f over an interval divided by the length of the interval." We make an analogous statement here: the average value of $z = f(x, y)$ over a region R is the volume under f over R divided by the area of R.

Definition 102 The Average Value of f on R

Let $z = f(x, y)$ be a continuous function defined over a closed region R in the xy-plane. The **average value of** f on R is

$$\text{average value of } f \text{ on } R = \frac{\iint_R f(x,y)\,dA}{\iint_R dA}. \tag{13.2.13}$$

Example 13.2.6: Finding average value of a function over a region R

Find the average value of $f(x,y)=4-y$ over the region R, which is bounded by the parabolas $y^2=4x$ and $x^2=4y$. Note: this is the same function and region as used in Example 13.2.4.

Solution

In Example 13.2.4 we found

$$\iint_R f(x,y)\,dA=\int_0^4\int_{y^2/4}^{2\sqrt{y}}(4-y)\,dx\,dy=\frac{176}{15}.$$

We find the area of R by computing $\iint_R dA$:

$$\iint_R dA=\int_0^4\int_{y^2/4}^{2\sqrt{y}}dx\,dy=\frac{16}{3}.$$

Dividing the volume under the surface by the area gives the average value:

$$\text{average value of } f \text{ on } R=\frac{176/15}{16/3}=\frac{11}{5}=2.2.$$

While the surface, as shown in Figure 13.2.9, covers z-values from $z=0$ to $z=4$, the "average" z-value on R is 2.2.

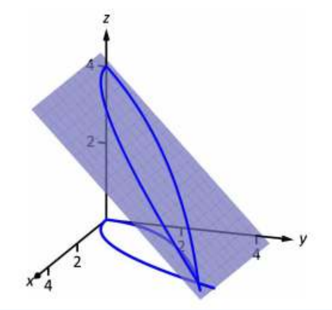


Figure 13.2.9: Finding the average value of f in Example 13.2.6.

The previous section introduced the iterated integral in the context of finding the area of plane regions. This section has extended our understanding of iterated integrals; now we see they can be used to find the signed volume under a surface.

This new understanding allows us to revisit what we did in the previous section. Given a region R in the plane, we computed $\iint_R 1\,dA$; again, our understanding at the time was that we were finding the area of R. However, we can now view the function $z=1$ as a surface, a flat surface with constant z-value of 1. The double integral $\iint_R 1\,dA$ finds the volume, under $z=1$, over R, as shown in Figure 13.2.10. Basic geometry tells us that if the base of a general right cylinder has area A, its volume is $A\cdot h$, where h is the height. In our case, the height is 1. We were "actually" computing the volume of a solid, though we interpreted the number as an area.

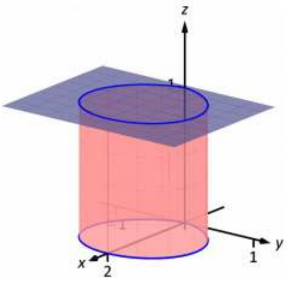


Figure 13.2.10: Showing how an iterated integral used to find area finds a certain volume.

The next section extends our abilities to find "volumes under surfaces." Currently, some integrals are hard to compute because either the region R we are integrating over is hard to define with rectangular curves, or the integrand itself is hard to deal with.

Some of these problems can be solved by converting everything into polar coordinates.

This page titled 13.2: Double Integration and Volume is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

13.3: Double Integration with Polar Coordinates

We have used iterated integrals to evaluate double integrals, which give the signed volume under a surface, $z = f(x, y)$, over a region R of the xy-plane. The integrand is simply $f(x, y)$, and the bounds of the integrals are determined by the region R.

Some regions R are easy to describe using rectangular coordinates -- that is, with equations of the form $y = f(x)$, $x = a$, etc. However, some regions are easier to handle if we represent their boundaries with polar equations of the form $r = f(\theta)$, $\theta = \alpha$, etc.

The basic form of the double integral is $\iint_R f(x, y)\, dA$. We interpret this integral as follows: over the region R, sum up lots of products of heights (given by $f(x_i, y_i)$) and areas (given by ΔA_i). That is, dA represents "a little bit of area." In rectangular coordinates, we can describe a small rectangle as having area $dx\, dy$ or $dy\, dx$ -- the area of a rectangle is simply length×width -- a small change in x times a small change in y. Thus we replace dA in the double integral with $dx\, dy$ or $dy\, dx$.

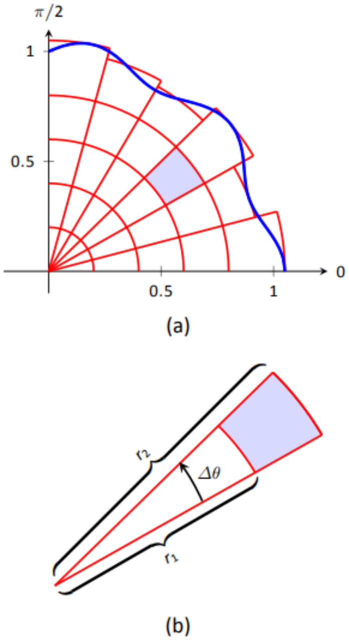


FIGURE 13.3.1

Now consider representing a region R with polar coordinates. Consider Figure $13.3.1a$. Let R be the region in the first quadrant bounded by the curve. We can approximate this region using the natural shape of polar coordinates: portions of sectors of circles. In the figure, one such region is shaded, shown again in part (b) of the figure.

As the area of a sector of a circle with radius r, subtended by an angle θ, is $A = \frac{1}{2}r^2\theta$, we can find the area of the shaded region. The whole sector has area $\frac{1}{2}r_2^2\Delta\theta$, whereas the smaller, unshaded sector has area $\frac{1}{2}r_1^2\Delta\theta$. The area of the shaded region is the difference of these areas:

$$\Delta A_i = \frac{1}{2}r_2^2\Delta\theta - \frac{1}{2}r_1^2\Delta\theta = \frac{1}{2}\left(r_2^2 - r_1^2\right)\left(\Delta\theta\right) = \frac{r_2 + r_1}{2}\left(r_2 - r_1\right)\Delta\theta. \tag{13.3.1}$$

Note that $(r_2 + r_1)/2$ is just the average of the two radii.

To approximate the region R, we use many such subregions; doing so shrinks the difference $r_2 - r_1$ between radii to 0 and shrinks the change in angle $\Delta\theta$ also to 0. We represent these infinitesimal changes in radius and angle as dr and $d\theta$, respectively. Finally, as dr is small, $r_2 \approx r_1$, and so $(r_2 + r_1)/2 \approx r_1$. Thus, when dr and $d\theta$ are small,

$$\Delta A_i \approx r_i\, dr\, d\theta. \tag{13.3.2}$$

Taking a limit, where the number of subregions goes to infinity and both $r_2 - r_1$ and $\Delta\theta$ go to 0, we get

$$dA = r\, dr\, d\theta. \tag{13.3.3}$$

So to evaluate $\iint_R f(x,y)\ dA$, replace dA with $r\,dr\,d\theta$. Convert the function $z=f(x,y)$ to a function with polar coordinates with the substitutions $x=r\cos\theta$, $y=r\sin\theta$. Finally, find bounds $g_1(\theta)\leq r\leq g_2(\theta)$ and $\alpha\leq\theta\leq\beta$ that describe R. This is the key principle of this section, so we restate it here as a Key Idea.

Key Idea: Evaluating Double Integrals with Polar Coordinates

Let R be a plane region bounded by the polar equations $\alpha\leq\theta\leq\beta$ and $g_1(\theta)\leq r\leq g_2(\theta)$. Then

$$\iint_R f(x,y)\ dA=\int_\alpha^\beta\int_{g_1(\theta)}^{g_2(\theta)} f\Big(r\cos\theta, r\sin\theta\Big)\ r\,dr\,d\theta. \tag{13.3.4}$$

Examples will help us understand this Key Idea.

Example 13.3.1: Evaluating a double integral with polar coordinates

Find the signed volume under the plane $z=4-x-2y$ over the circle with equation $x^2+y^2=1$.

Solution

The bounds of the integral are determined solely by the region R over which we are integrating. In this case, it is a circle with equation $x^2+y^2=1$. We need to find polar bounds for this region. It may help to review polar coordinates earlier in this text; bounds for this circle are $0\leq r\leq 1$ and $0\leq\theta\leq 2\pi$.

We replace $f(x,y)$ with $f(r\cos\theta, r\sin\theta)$. That means we make the following substitutions:

$$4-x-2y \quad\Rightarrow\quad 4-r\cos\theta-2r\sin\theta.$$

Finally, we replace dA in the double integral with $r\,dr\,d\theta$. This gives the final iterated integral, which we evaluate:

$$\begin{aligned}\iint_R f(x,y)\ dA &= \int_0^{2\pi}\int_0^1\Big(4-r\cos\theta-2r\sin\theta\Big)r\,dr\,d\theta\\ &=\int_0^{2\pi}\int_0^1\Big(4r-r^2(\cos\theta-2\sin\theta)\Big)dr\,d\theta\\ &=\int_0^{2\pi}\left(2r^2-\frac{1}{3}r^3(\cos\theta-2\sin\theta)\right)\Bigg|_0^1\,d\theta\\ &=\int_0^{2\pi}\left(2-\frac{1}{3}\Big(\cos\theta-2\sin\theta\Big)\right)\,d\theta\\ &=\left(2\theta-\frac{1}{3}\Big(\sin\theta+2\cos\theta\Big)\right)\Bigg|_0^{2\pi}\\ &=4\pi\approx 12.566.\end{aligned}$$

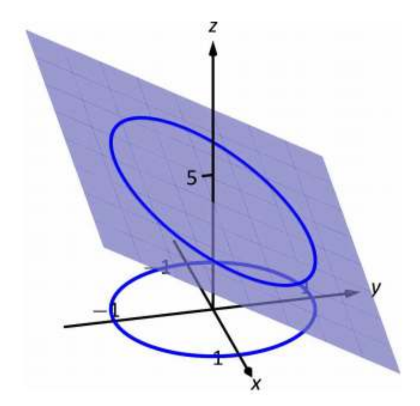


FIGURE 13.3.2

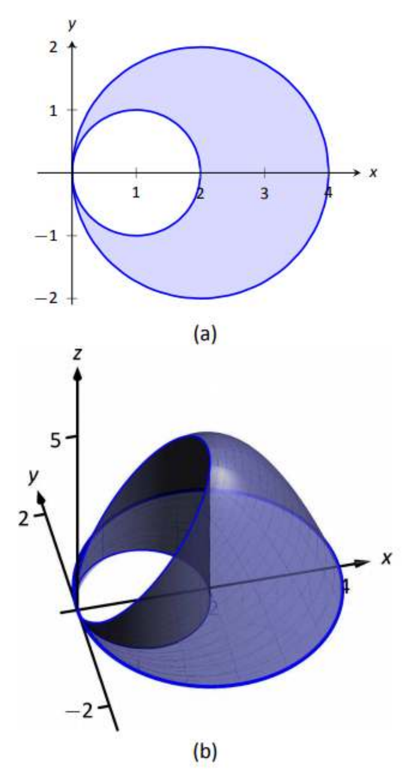
The surface and region R are shown in Figure 13.3.2.

Example $13.3.2$: Evaluating a double integral with polar coordinates

Find the volume under the paraboloid $z=4-(x-2)^2-y^2$ over the region bounded by the circles $(x-1)^2+y^2=1$ and $(x-2)^2+y^2=4$.

Solution

At first glance, this seems like a very hard volume to compute as the region R (shown in Figure $13.3.3a$) has a hole in it, cutting out a strange portion of the surface, as shown in part (b) of the figure. However, by describing R in terms of polar equations, the volume is not very difficult to compute.



(a)

(b)

FIGURE $13.3.3$

It is straightforward to show that the circle $(x-1)^2+y^2=1$ has polar equation $r=2\cos\theta$, and that the circle $(x-2)^2+y^2=4$ has polar equation $r=4\cos\theta$. Each of these circles is traced out on the interval $0\le\theta\le\pi$. The bounds on r are $2\cos\theta\le r\le 4\cos\theta$.

Replacing x with $r\cos\theta$ in the integrand, along with replacing y with $r\sin\theta$, prepares us to evaluate the double integral $\iint_R f(x,y)\,dA$:

$$\begin{aligned}\iint_R f(x,y)\,dA &= \int_0^{\pi}\int_{2\cos\theta}^{4\cos\theta}\Big(4-\big(r\cos\theta-2\big)^2-\big(r\sin\theta\big)^2\Big)r\,dr\,d\theta\\ &= \int_0^{\pi}\int_{2\cos\theta}^{4\cos\theta}\Big(-r^3+4r^2\cos\theta\Big)dr\,d\theta\\ &= \int_0^{\pi}\Big(-\frac{1}{4}r^4+\frac{4}{3}r^3\cos\theta\Big)\Big|_{2\cos\theta}^{4\cos\theta}d\theta\\ &= \int_0^{\pi}\Big(\Big[-\frac{1}{4}(256\cos^4\theta)+\frac{4}{3}(64\cos^4\theta)\Big]-\\ &\quad\Big[-\frac{1}{4}(16\cos^4\theta)+\frac{4}{3}(8\cos^4\theta)\Big]\Big)\,d\theta\\ &= \int_0^{\pi}\frac{44}{3}\cos^4\theta\,d\theta.\end{aligned}$$

To integrate $\cos^4\theta$, rewrite it as $\cos^2\theta\cos^2\theta$ and employ the power-reducing formula twice:

$$\begin{aligned}\cos^4\theta &= \cos^2\theta\cos^2\theta \\ &= \frac{1}{2}\big(1+\cos(2\theta)\big)\frac{1}{2}\big(1+\cos(2\theta)\big) \\ &= \frac{1}{4}\big(1+2\cos(2\theta)+\cos^2(2\theta)\big) \\ &= \frac{1}{4}\Big(1+2\cos(2\theta)+\frac{1}{2}\big(1+\cos(4\theta)\big)\Big) \\ &= \frac{3}{8}+\frac{1}{2}\cos(2\theta)+\frac{1}{8}\cos(4\theta).\end{aligned}$$

Picking up from where we left off above, we have

$$\begin{aligned}&= \int_0^\pi \frac{44}{3}\cos^4\theta\,d\theta \\ &= \int_0^\pi \frac{44}{3}\left(\frac{3}{8}+\frac{1}{2}\cos(2\theta)+\frac{1}{8}\cos(4\theta)\right)d\theta \\ &= \frac{44}{3}\left(\frac{3}{8}\theta+\frac{1}{4}\sin(2\theta)+\frac{1}{32}\sin(4\theta)\right)\Big|_0^\pi \\ &= \frac{11}{2}\pi \approx 17.279.\end{aligned}$$

While this example was not trivial, the double integral would have been *much* harder to evaluate had we used rectangular coordinates.

Example 13.3.3: Evaluating a double integral with polar coordinates

Find the volume under the surface $f(x,y)=\frac{1}{x^2+y^2+1}$ over the sector of the circle with radius a centered at the origin in the first quadrant, as shown in Figure 13.3.4.

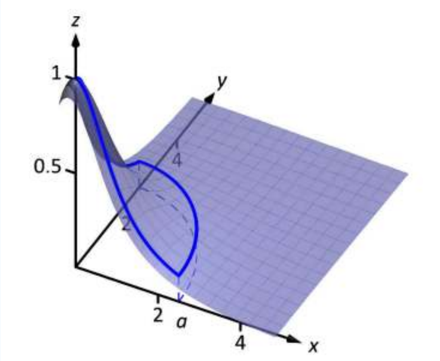


FIGURE 13.3.4

Solution

The region R we are integrating over is a circle with radius a, restricted to the first quadrant. Thus, in polar, the bounds on R are $0\le r\le a$, $0\le\theta\le\pi/2$. The integrand is rewritten in polar as

$$\frac{1}{x^2+y^2+1} \Rightarrow \frac{1}{r^2\cos^2\theta+r^2\sin^2\theta+1}=\frac{1}{r^2+1}. \tag{13.3.5}$$

We find the volume as follows:

$$\begin{aligned}\iint_R f(x,y)\,dA &= \int_0^{\pi/2}\int_0^a \frac{r}{r^2+1}\,dr\,d\theta \\ &= \int_0^{\pi/2} \frac{1}{2}\Big(\ln\left|r^2+1\right|\Big)\Big|_0^a\,d\theta \\ &= \int_0^{\pi/2}\frac{1}{2}\ln(a^2+1)\,d\theta \\ &= \left(\frac{1}{2}\ln(a^2+1)\theta\right)\Big|_0^{\pi/2} \\ &= \frac{\pi}{4}\ln(a^2+1).\end{aligned}$$

Figure $13.3.4$ shows that f shrinks to near 0 very quickly. Regardless, as a grows, so does the volume, without bound.

Note: Previous work has shown that there is finite *area* under $\frac{1}{x^2+1}$ over the entire x-axis. However, Example $13.3.3$ shows that there is infinite *volume* under $\dfrac{1}{x^2+y^2+1}$ over the entire xy-plane.

Example $13.3.4$: Finding the volume of a sphere

Find the volume of a sphere with radius a.

Solution

The sphere of radius a, centered at the origin, has equation $x^2+y^2+z^2=a^2$; solving for z, we have $z=\sqrt{a^2-x^2-y^2}$. This gives the upper half of a sphere. We wish to find the volume under this top half, then double it to find the total volume.

The region we need to integrate over is the circle of radius a, centered at the origin. Polar bounds for this equation are $0\le r\le a$, $0\le\theta\le 2\pi$.

All together, the volume of a sphere with radius a is:

$$\begin{aligned}2\iint_R\sqrt{a^2-x^2-y^2}\,dA &= 2\int_0^{2\pi}\int_0^a\sqrt{a^2-(r\cos\theta)^2-(r\sin\theta)^2}\,r\,dr\,d\theta \\ &= 2\int_0^{2\pi}\int_0^a r\sqrt{a^2-r^2}\,dr\,d\theta.\end{aligned}\tag{13.3.6}$$

We can evaluate this inner integral with substitution. With $u=a^2-r^2$, $du=-2r\,dr$. The new bounds of integration are $u(0)=a^2$ to $u(a)=0$. Thus we have:

$$\begin{aligned}&= \int_0^{2\pi}\int_{a^2}^0\left(-u^{1/2}\right)du\,d\theta \\ &= \int_0^{2\pi}\left(-\frac{2}{3}u^{3/2}\right)\Big|_{a^2}^0\,d\theta \\ &= \int_0^{2\pi}\left(\frac{2}{3}a^3\right)d\theta \\ &= \left(\frac{2}{3}a^3\theta\right)\Big|_0^{2\pi} \\ &= \frac{4}{3}\pi a^3.\end{aligned}$$

Generally, the formula for the volume of a sphere with radius r is given as $4/3\pi r^3$; we have justified this formula with our calculation.

Example $13.3.5$: Finding the volume of a solid

A sculptor wants to make a solid bronze cast of the solid shown in Figure $13.3.5$, where the base of the solid has boundary, in polar coordinates, $r=\cos(3\theta)$, and the top is defined by the plane $z=1-x+0.1y$. Find the volume of the solid.

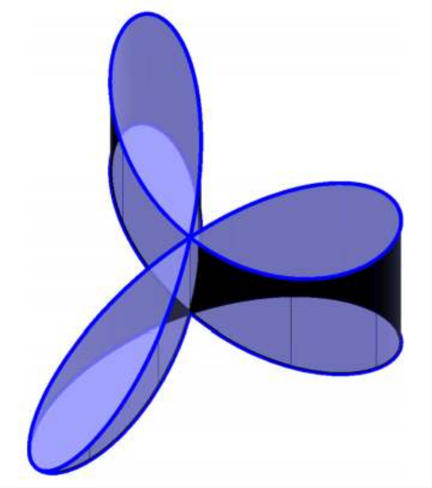


FIGURE 13.3.5

Solution

From the outset, we should recognize that knowing *how to set up* this problem is probably more important than knowing *how to compute the integrals*. The iterated integral to come is not "hard" to evaluate, though it is long, requiring lots of algebra. Once the proper iterated integral is determined, one can use readily--available technology to help compute the final answer.

The region R that we are integrating over is bound by $0 \le r \le \cos(3\theta)$, for $0 \le \theta \le \pi$ (note that this rose curve is traced out on the interval $[0,\pi]$, not $[0,2\pi]$). This gives us our bounds of integration. The integrand is $z = 1 - x + 0.1y$; converting to polar, we have that the volume V is:

$$V = \iint_R f(x,y)\ dA = \int_0^\pi \int_0^{\cos(3\theta)} \big(1 - r\cos\theta + 0.1r\sin\theta\big) r\,dr\,d\theta.$$

Distributing the r, the inner integral is easy to evaluate, leading to

$$\int_0^\pi \left(\frac{1}{2}\cos^2(3\theta) - \frac{1}{3}\cos^3(3\theta)\cos\theta + \frac{0.1}{3}\cos^3(3\theta)\sin\theta\right)\ d\theta.$$

This integral takes time to compute by hand; it is rather long and cumbersome. The powers of cosine need to be reduced, and products like $\cos(3\theta)\cos\theta$ need to be turned to sums using the Product To Sum formulas in the back cover of this text.

We rewrite $\frac{1}{2}\cos^2(3\theta)$ as $\frac{1}{4}(1+\cos(6\theta))$. We can also rewrite $\frac{1}{3}\cos^3(3\theta)\cos\theta$ as:

$$\frac{1}{3}\cos^3(3\theta)\cos\theta = \frac{1}{3}\cos^2(3\theta)\cos(3\theta)\cos\theta = \frac{1}{3}\frac{1+\cos(6\theta)}{2}\big(\cos(4\theta) + \cos(2\theta)\big).$$

This last expression still needs simplification, but eventually all terms can be reduced to the form $a\cos(m\theta)$ or $a\sin(m\theta)$ for various values of a and m.

We forgo the algebra and recommend the reader employ technology, such as WolframAlpha, to compute the numeric answer. Such technology gives:

$$\int_0^\pi \int_0^{\cos(3\theta)} \big(1 - r\cos\theta + 0.1r\sin\theta\big) r\,dr\,d\theta = \frac{\pi}{4} \approx 0.785u^3.$$

Since the units were not specified, we leave the result as almost 0.8 cubic units (meters, feet, etc.) Should the artist want to scale the piece uniformly, so that each rose petal had a length other than 1, she should keep in mind that scaling by a factor of k scales the volume by a factor of k^3.

We have used iterated integrals to find areas of plane regions and volumes under surfaces. Just as a single integral can be used to compute much more than "area under the curve," iterated integrals can be used to compute much more than we have thus far seen. The next two sections show two, among many, applications of iterated integrals.

This page titled 13.3: Double Integration with Polar Coordinates is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

13.4: Center of Mass

We have used iterated integrals to find areas of plane regions and signed volumes under surfaces. A brief recap of these uses will be useful in this section as we apply iterated integrals to compute the **mass** and **center of mass** of planar regions.

To find the area of a planar region, we evaluated the double integral $\iint_R dA$. That is, summing up the areas of lots of little subregions of R gave us the total area. Informally, we think of $\iint_R dA$ as meaning "sum up lots of little areas over R."

To find the signed volume under a surface, we evaluated the double integral $\iint_R f(x,y)\,dA$. Recall that the "dA" is not just a "bookend" at the end of an integral; rather, it is multiplied by $f(x,y)$. We regard $f(x,y)$ as giving a height, and dA still giving an area: $f(x,y)\,dA$ gives a volume. Thus, informally, $\iint_R f(x,y)\,dA$ means "sum up lots of little volumes over R."

We now extend these ideas to other contexts.

Mass and Weight

Consider a thin sheet of material with constant thickness and finite area. Mathematicians (and physicists and engineers) call such a sheet a **lamina**. So consider a lamina, as shown in Figure $13.4.1a$, with the shape of some planar region R, as shown in part (b).

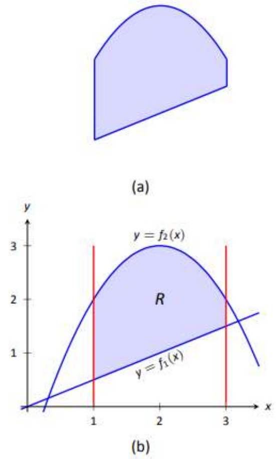


Figure $13.4.1$: Illustrating the concept of a lamina.

We can write a simple double integral that represents the mass of the lamina: $\iint_R dm$, where "dm" means "a little mass." That is, the double integral states the total mass of the lamina can be found by "summing up lots of little masses over R."

To evaluate this double integral, partition R into n subregions as we have done in the past. The i^{th} subregion has area ΔA_i.
A fundamental property of mass is that "mass=density×area." If the lamina has a constant density δ, then the mass of this i^{th} subregion is $\Delta m_i = \delta\Delta A_i$. That is, we can compute a small amount of mass by multiplying a small amount of area by the density.

If density is variable, with density function $\delta = \delta(x,y)$, then we can approximate the mass of the i^{th} subregion of R by multiplying ΔA_i by $\delta(x_i, y_i)$, where (x_i, y_i) is a point in that subregion. That is, for a small enough subregion of R, the density across that region is almost constant.

Note: *Mass* and *weight* are different measures. Since they are scalar multiples of each other, it is often easy to treat them as the same measure. In this section we effectively treat them as the same, as our technique for finding mass is the same as for finding weight. The density functions used will simply have different units.

The total mass M of the lamina is approximately the sum of approximate masses of subregions:

$$M \approx \sum_{i=1}^{n} \Delta m_i = \sum_{i=1}^{n} \delta(x_i, y_i)\,\Delta A_i.$$

Taking the limit as the size of the subregions shrinks to 0 gives us the actual mass; that is, integrating $\delta(x,y)$ over R gives the mass of the lamina.

Definition 103 Mass of a Lamina with VarIable Density

Let $\delta(x,y)$ be a continuous density function of a lamina corresponding to a plane region R. The mass M of the lamina is

$$\text{mass } M=\iint_R\,dm=\iint_R\delta(x,y)\,dA. \tag{13.4.1}$$

Example $13.4.1$: Finding the mass of a lamina with constant density

Find the mass of a square lamina, with side length 1, with a density of $\delta=3$ gm/cm².

Solution

We represent the lamina with a square region in the plane as shown in Figure $13.4.2$. As the density is constant, it does not matter where we place the square.

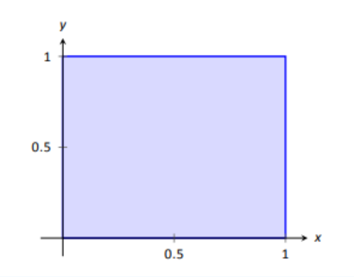


Figure $13.4.2$: A region R representing a lamina in Example $13.4.1$.

Following Definition 103, the mass M of the lamina is

$$M=\iint_R 3\,dA=\int_0^1\int_0^1 3\,dx\,dy=3\int_0^1\int_0^1 dx\,dy=3\text{ gm}.$$

This is all very straightforward; note that all we really did was find the area of the lamina and multiply it by the constant density of 3 gm/cm^2.

Example $13.4.2$: Finding the mass of a lamina with variable density

Find the mass of a square lamina, represented by the unit square with lower lefthand corner at the origin (see Figure $13.4.2$), with variable density $\delta(x,y)=(x+y+2)$ gm/cm^2.

Solution

The variable density δ, in this example, is very uniform, giving a density of 3 in the center of the square and changing linearly. A graph of $\delta(x,y)$ can be seen in Figure $13.4.3$; notice how "same amount" of density is above $z=3$ as below. We'll comment on the significance of this momentarily.

The mass M is found by integrating $\delta(x,y)$ over R. The order of integration is not important; we choose $dx\,dy$ arbitrarily. Thus:

$$\begin{aligned}M=\iint_R(x+y+2)dA&=\int_0^1\int_0^1(x+y+2)\ dx\,dy\\&=\int_0^1\left(\frac{1}{2}x^2+x(y+2)\right)\Big|_0^1\ dy\\&=\int_0^1\left(\frac{5}{2}+y\right)\ dy\\&=\left(\frac{5}{2}y+\frac{1}{2}y^2\right)\Big|_0^1\\&=3\text{ gm}.\end{aligned}$$

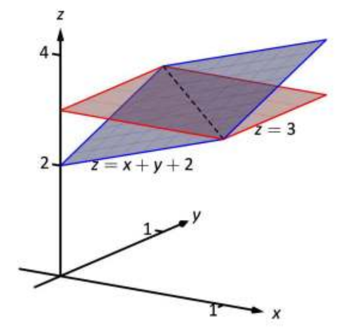


Figure $13.4.3$: Graphing the density functions in Examples $13.4.1$ and $13.4.2$.

It turns out that since since the density of the lamina is so uniformly distributed "above and below" $z = 3$ that the mass of the lamina is the same as if it had a constant density of 3. The density functions in Examples $13.4.1$ and $13.4.2$ are graphed in Figure $13.4.3$, which illustrates this concept.

Example $13.4.3$: Finding the weight of a lamina with variable density

Find the weight of the lamina represented by the circle with radius 2 ft, centered at the origin, with density function $\delta(x,y) = (x^2 + y^2 + 1)$ lb/ft^2. Compare this to the weight of the same lamina with density $\delta(x,y) = (2\sqrt{x^2+y^2} + 1)$ lb/ft^2.

Solution

A direct application of Definition 103 states that the weight of the lamina is $\iint_R \delta(x,y)\, dA$. Since our lamina is in the shape of a circle, it makes sense to approach the double integral using polar coordinates.

The density function $\delta(x,y) = x^2 + y^2 + 1$ becomes $\delta(r,\theta) = (r\cos\theta)^2 + (r\sin\theta)^2 + 1 = r^2 + 1$. The circle is bounded by $0 \le r \le 2$ and $0 \le \theta \le 2\pi$. Thus the weight W is:

$$\begin{aligned} W &= \int_0^{2\pi}\int_0^2 (r^2+1)r\ dr\ d\theta \\ &= \int_0^{2\pi} \left(\frac{1}{4}r^4 + \frac{1}{2}r^2\right)\Big|_0^2\ d\theta \\ &= \int_0^{2\pi} (6)\ d\theta \\ &= 12\pi \approx 37.70 \text{ lb.} \end{aligned}$$

Now compare this with the density function $\delta(x,y) = 2\sqrt{x^2+y^2} + 1$. Converting this to polar coordinates gives $\delta(r,\theta) = 2\sqrt{(r\cos\theta)^2 + (r\sin\theta)^2} + 1 = 2r + 1$. Thus the weight W is:

$$\begin{aligned} W &= \int_0^{2\pi}\int_0^2 (2r+1)r\ dr\ d\theta \\ &= \int_0^{2\pi} (\frac{2}{3}r^3 + \frac{1}{2}r^2)\Big|_0^2 d\theta \\ &= \int_0^{2\pi} \left(\frac{22}{3}\right)\ d\theta \\ &= \frac{44}{3}\pi \approx 46.08 \text{ lb.} \end{aligned}$$

One would expect different density functions to return different weights, as we have here. The density functions were chosen, though, to be similar: each gives a density of 1 at the origin and a density of 5 at the outside edge of the circle, as seen in Figure $13.4.4$.

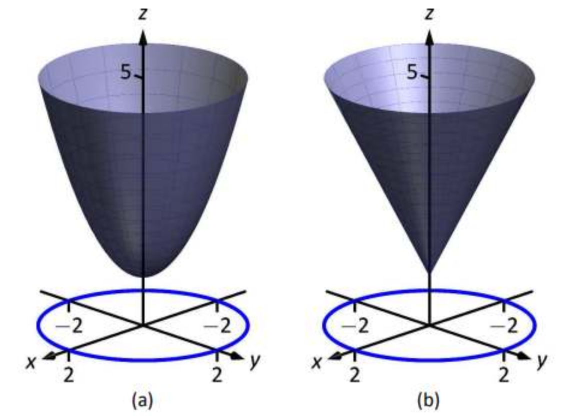


Figure 13.4.4: Graphing the density functions in 13.4.3. In (a) is the density function $\delta(x,y)=x^2+y^2+1$; in (b) is $\delta(x,y)=2\sqrt{x^2+y^2}+1$.

Notice how $x^2+y^2+1\leq 2\sqrt{x^2+y^2}+1$ over the circle; this results in less weight.

Plotting the density functions can be useful as our understanding of mass can be related to our understanding of "volume under a surface." We interpreted $\iint_R f(x,y)\,dA$ as giving the volume under f over R; we can understand $\iint_R \delta(x,y)\,dA$ in the same way. The "volume" under δ over R is actually mass; by compressing the "volume" under δ onto the xy-plane, we get "more mass" in some areas than others -- i.e., areas of greater density.

Knowing the mass of a lamina is one of several important measures. Another is the **center of mass**, which we discuss next.

Center of Mass

Consider a disk of radius 1 with uniform density. It is common knowledge that the disk will balance on a point if the point is placed at the center of the disk. What if the disk does not have a uniform density? Through trial-and-error, we should still be able to find a spot on the disk at which the disk will balance on a point. This balance point is referred to as the **center of mass**, or **center of gravity**. It is though all the mass is "centered" there. In fact, if the disk has a mass of 3 kg, the disk will behave physically as though it were a point-mass of 3 kg located at its center of mass. For instance, the disk will naturally spin with an axis through its center of mass (which is why it is important to "balance" the tires of your car: if they are "out of balance", their center of mass will be outside of the axle and it will shake terribly).

We find the center of mass based on the principle of a **weighted average**. Consider a college class in which your homework average is 90%, your test average is 73%, and your final exam grade is an 85%. Experience tells us that our final grade *is not* the *average* of these three grades: that is, it is not:

$$\frac{0.9+0.73+0.85}{3}\approx 0.837=83.7 \tag{13.4.2}$$

That is, you are probably not pulling a B in the course. Rather, your grades are *weighted*. Let's say the homework is worth 10% of the grade, tests are 60% and the exam is 30%. Then your final grade is:

$$(0.1)(0.9)+(0.6)(0.73)+(0.3)(0.85)=0.783=78.3 \tag{13.4.3}$$

Each grade is multiplied by a **weight**.

In general, given values $x_1,x_2,\ldots,x_n$ and weights $w_1,w_2,\ldots,w_n$, the weighted average of the n values is

$$\sum_{i=1}^n w_ix_i\Big/\sum_{i=1}^n w_i.$$

In the grading example above, the sum of the weights 0.1, 0.6 and 0.3 is 1, so we don't see the division by the sum of weights in that instance.

How this relates to center of mass is given in the following theorem.

THEOREM 121 Center of Mass of Discrete Linear System

Let point masses $m_1, m_2, \ldots, m_n$ be distributed along the x-axis at locations $x_1, x_2, \ldots, x_n$, respectively. The center of mass $\overline{x}$ of the system is located at

$$\overline{x} = \sum_{i=1}^{n} m_i x_i \Bigg/ \sum_{i=1}^{n} m_i. \tag{13.4.4}$$

Example 13.4.4: Finding the center of mass of a discrete linear system

1. Point masses of 2 gm are located at $x=-1$, $x=2$ and $x=3$ are connected by a thin rod of negligible weight. Find the center of mass of the system.
2. Point masses of 10 gm, 2 gm and 1 gm are located at $x=-1$, $x=2$ and $x=3$, respectively, are connected by a thin rod of negligible weight. Find the center of mass of the system.

Solution

1. Following Theorem 121, we compute the center of mass as:

$$\overline{x} = \frac{2(-1)+2(2)+2(3)}{2+2+2} = \frac{4}{3} = 1.\overline{3}. \tag{13.4.5}$$

So the system would balance on a point placed at $x = 4/3$, as illustrated in Figure $13.4.5a$.

2. Again following Theorem 121, we find:

$$\overline{x} = \frac{10(-1)+2(2)+1(3)}{10+2+1} = \frac{-3}{13} \approx -0.23. \tag{13.4.6}$$

Placing a large weight at the left hand side of the system moves the center of mass left, as shown in Figure $13.4.5b$.

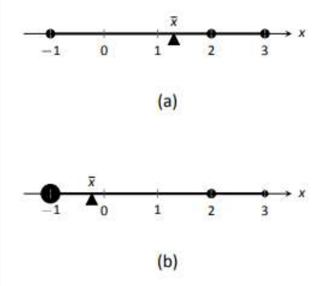


Figure $13.4.5$: Illustrating point masses along a thin rod and the center of mass.

In a discrete system (i.e., mass is located at individual points, not along a continuum) we find the center of mass by dividing the mass into a **moment** of the system. In general, a moment is a weighted measure of distance from a particular point or line. In the case described by Theorem 121, we are finding a weighted measure of distances from the y-axis, so we refer to this as **the moment about the y-axis**, represented by M_y. Letting M be the total mass of the system, we have $\overline{x} = M_y/M$.

We can extend the concept of the center of mass of discrete points along a line to the center of mass of discrete points in the plane rather easily. To do so, we define some terms then give a theorem.

Definition 104 Moments about the x- and y- Axes.

Let point masses m_1, $m_2, \ldots, m_n$ be located at points (x_1, y_1), $(x_2, y_2) \ldots, (x_n, y_n)$, respectively, in the xy-plane.

1. The **moment about the** y-axis, M_y, is $M_y = \sum_{i=1}^{n} m_i x_i$.
2. The **moment about the** x-axis}, M_x, is $M_x = \sum_{i=1}^{n} m_i y_i$.

One can think that these definitions are "backwards" as M_y sums up "x" distances. But remember, "x" distances are measurements of distance from the y-axis, hence defining the moment about the y-axis.

We now define the center of mass of discrete points in the plane.

THEOREM 122 Center of Mass of Discrete Planar System

Let point masses $m_1,\, m_2, \ldots, m_n$ be located at points $(x_1, y_1),\, (x_2, y_2) \ldots, (x_n, y_n)$, respectively, in the xy-plane, and let $M = \sum_{i=1}^{n} m_i$.

The center of mass of the system is at $(\overline{x},\, \overline{y})$, where

$$\overline{x} = \frac{M_y}{M} \quad \text{and} \quad \overline{y} = \frac{M_x}{M}. \tag{13.4.7}$$

Example 13.4.5: Finding the center of mass of a discrete planar system

Let point masses of 1 kg, 2 kg and 5 kg be located at points $(2, 0)$, $(1, 1)$ and $(3, 1)$, respectively, and are connected by thin rods of negligible weight. Find the center of mass of the system.

Solution

We follow Theorem 122 and Definition 104 to find M, M_x and M_y:

$M = 1 + 2 + 5 = 8$kg.

$$\begin{aligned} M_x &= \sum_{i=1}^{n} m_i y_i \\ &= 1(0) + 2(1) + 5(1) \\ &= 7. \end{aligned}$$

$$\begin{aligned} M_y &= \sum_{i=1}^{n} m_i x_i \\ &= 1(2) + 2(1) + 5(3) \\ &= 19. \end{aligned}$$

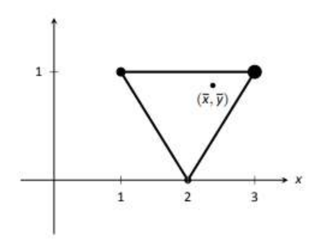


Figure $13.4.6$: Illustrating the center of mass of a discrete planar system in Example $13.4.5$.

Thus the center of mass is $(\overline{x},\, \overline{y}) = \left(\frac{M_y}{M}, \frac{M_x}{M}\right) = \left(\frac{1}{9}8, \frac{7}{8}\right) = (2.375, 0.875)$, illustrated in Figure $13.4.6$.

We finally arrive at our true goal of this section: finding the center of mass of a lamina with variable density. While the above measurement of center of mass is interesting, it does not directly answer more realistic situations where we need to find the center of mass of a contiguous region. However, understanding the discrete case allows us to approximate the center of mass of a planar lamina; using calculus, we can refine the approximation to an exact value.

We begin by representing a planar lamina with a region R in the xy-plane with density function $\delta(x, y)$. Partition R into n subdivisions, each with area ΔA_i. As done before, we can approximate the mass of the i^{th} subregion with $\delta(x_i,\, y_i)\, \Delta A_i$, where $(x_i,\, y_i)$ is a point inside the i^{th} subregion. We can approximate the moment of this subregion about the y-axis with $x_i\, \delta(x_i, y_i)\, \Delta A_i$ -- that is, by multiplying the approximate mass of the region by its approximate distance from the y-axis. Similarly, we can approximate the moment about the x-axis with $y_i\, \delta(x_i, y_i)\, \Delta A_i$. By summing over all subregions, we have:

$$\text{mass: } M \approx \sum_{i=1}^{n} \delta(x_i, y_i)\,\Delta A_i \quad \text{(as seen before)}$$
$$\text{moment about the } x\text{-axis: } M_x \approx \sum_{i=1}^{n} y_i\delta(x_i, y_i)\,\Delta A_i$$
$$\text{moment about the } y\text{-axis: } M_y \approx \sum_{i=1}^{n} x_i\delta(x_i, y_i)\,\Delta A_i$$

By taking limits, where size of each subregion shrinks to 0 in both the x and y directions, we arrive at the double integrals given in the following theorem.

Theorem 123: Center of Mass of a Planar Lamina, Moments

Let a planar lamina be represented by a region R in the xy-plane with density function $\delta(x, y)$.

1. mass: $M = \iint_R \delta(x, y)\, dA$
2. moment about the x-axis: $M_x = \iint_R y\delta(x, y)\, dA$
3. moment about the y-axis: $M_y = \iint_R x\delta(x, y)\, dA$
4. The center of mass of the lamina is

$$(\overline{x}, \overline{y}) = \left(\frac{M_y}{M}, \frac{M_x}{M}\right).$$

We start our practice of finding centers of mass by revisiting some of the lamina used previously in this section when finding mass. We will just set up the integrals needed to compute M, M_x and M_y and leave the details of the integration to the reader.

Example 13.4.6: Finding the center of mass of a lamina

Find the center mass of a square lamina, with side length 1, with a density of $\delta = 3$ gm/cm². (Note: this is the lamina from Example 13.4.1.)

Solution

We represent the lamina with a square region in the plane as shown in Figure 13.4.7 as done previously.

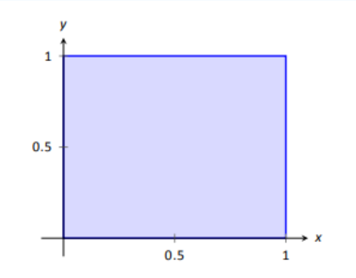


Figure 13.4.7: A region R representing a lamina in Example 13.4.6.

Following Theorem 123, we find M, M_x and M_y:

$$M = \iint_R 3\, dA = \int_0^1 \int_0^1 3\, dx\, dy = 3 \text{ gm}.$$
$$M_x = \iint_R 3y\, dA = \int_0^1 \int_0^1 3y\, dx\, dy = 3/2 = 1.5.$$
$$M_y = \iint_R 3x\, dA = \int_0^1 \int_0^1 3x\, dx\, dy = 3/2 = 1.5.$$

Thus the center of mass is $(\overline{x}, \overline{y}) = \left(\frac{M_y}{M}, \frac{M_x}{M}\right) = (1.5/3, 1.5/3) = (0.5, 0.5)$. This is what we should have expected: the center of mass of a square with constant density is the center of the square.

Example 13.4.7: Finding the center of mass of a lamina

Find the center of mass of a square lamina, represented by the unit square with lower left-hand corner at the origin (see Figure 13.4.7), with variable density $\delta(x,y)=(x+y+2)$ gm/cm². (Note: this is the lamina from Example 13.4.2.)

Solution

We follow Theorem 123, to find M, M_x and M_y:

$$\begin{aligned} M &= \iint_R (x+y+2)\,dA = \int_0^1\int_0^1 (x+y+2)\ dx\,dy = 3 \text{ gm}. \\ M_x &= \iint_R y(x+y+2)\,dA = \int_0^1\int_0^1 y(x+y+2)\ dx\,dy = \frac{19}{12}. \\ M_y &= \iint_R x(x+y+2)\,dA = \int_0^1\int_0^1 x(x+y+2)\ dx\,dy = \frac{19}{12}. \end{aligned}$$

Thus the center of mass is $(\overline{x},\overline{y})=\left(\frac{M_y}{M},\frac{M_x}{M}\right)=\left(\frac{19}{36},\frac{19}{36}\right)\approx(0.528,0.528)$. While the mass of this lamina is the same as the lamina in the previous example, the greater density found with greater x and y values pulls the center of mass from the center slightly towards the upper right-hand corner.

Example 13.4.8: Finding the center of mass of a lamina

Find the center of mass of the lamina represented by the circle with radius 2 ft, centered at the origin, with density function $\delta(x,y)=(x^2+y^2+1)$ lb/ft². (Note: this is one of the lamina used in Example 13.4.3.)

Solution

As done in Example 13.4.3, it is best to describe R using polar coordinates.
Thus when we compute M_y, we will integrate not $x\delta(x,y)=x(x^2+y^2+1)$, but rather $\big(r\cos\theta\big)\delta(r\cos\theta,r\sin\theta)=\big(r\cos\theta\big)\big(r^2+1\big)$. We compute M, M_x and M_y:

$$\begin{aligned} M &= \int_0^{2\pi}\int_0^2 (r^2+1)r\ dr\ d\theta = 12\pi\approx 37.7 \text{ lb}. \\ M_x &= \int_0^{2\pi}\int_0^2 (r\sin\theta)(r^2+1)r\ dr\ d\theta = 0. \\ M_y &= \int_0^{2\pi}\int_0^2 (r\cos\theta)(r^2+1)r\ dr\ d\theta = 0. \end{aligned}$$

Since R and the density of R are both symmetric about the x and y axes, it should come as no big surprise that the moments about each axis is 0. Thus the center of mass is $(\overline{x},\overline{y})=(0,0)$.

Example 13.4.9: Finding the center of mass of a lamina

Find the center of mass of the lamina represented by the region R shown in Figure 13.4.8, half an annulus with outer radius 6 and inner radius 5, with constant density 2 lb/ft².

Solution

Once again it will be useful to represent R in polar coordinates. Using the description of R and/or the illustration, we see that R is bounded by $5\le r\le 6$ and $0\le\theta\le\pi$. As the lamina is symmetric about the y-axis, we should expect $M_y=0$. We compute M, M_x and M_y:

$$\begin{aligned} M &= \int_0^{\pi}\int_5^6 (2)r\ dr\ d\theta = 11\pi \text{ lb}. \\ M_x &= \int_0^{\pi}\int_5^6 (r\sin\theta)(2)r\ dr\ d\theta = \frac{364}{3}\approx 121.33. \\ M_y &= \int_0^{\pi}\int_5^6 (r\cos\theta)(2)r\ dr\ d\theta = 0. \end{aligned}$$

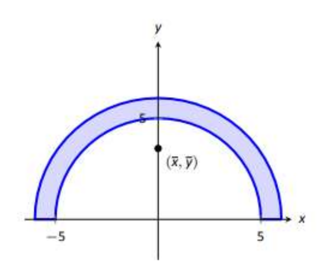


Figure 13.4.8: Illustrating the region R in Example 13.4.9.

Thus the center of mass is $(\overline{x}, \overline{y}) = \left(0, \frac{364}{33\pi}\right) \approx (0,\ 3.51)$. The center of mass is indicated in Figure 13.4.8; note how it lies outside of R!

This section has shown us another use for iterated integrals beyond finding area or signed volume under the curve. While there are many uses for iterated integrals, we give one more application in the following section: computing surface area.

This page titled 13.4: Center of Mass is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

13.5: Surface Area

In Section 7.4 we used definite integrals to compute the arc length of plane curves of the form $y = f(x)$. We later extended these ideas to compute the arc length of plane curves defined by parametric or polar equations.

The natural extension of the concept of "arc length over an interval" to surfaces is "surface area over a region."

Consider the surface $z = f(x, y)$ over a region R in the xy-plane, shown in Figure $13.5.1a$. Because of the domed shape of the surface, the surface area will be greater than that of the area of the region R. We can find this area using the same basic technique we have used over and over: we'll make an approximation, then using limits, we'll refine the approximation to the exact value.

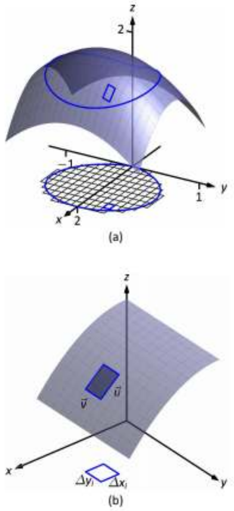


Figure $13.5.1$: Developing a method of computing surface area.

As done to find the volume under a surface or the mass of a lamina, we subdivide R into n subregions. Here we subdivide R into rectangles, as shown in the figure. One such subregion is outlined in the figure, where the rectangle has dimensions Δx_i and Δy_i, along with its corresponding region on the surface.

In part (b) of the figure, we zoom in on this portion of the surface. When Δx_i and Δy_i are small, the function is approximated well by the tangent plane at any point (x_i, y_i) in this subregion, which is graphed in part (b). In fact, the tangent plane approximates the function so well that in this figure, it is virtually indistinguishable from the surface itself! Therefore we can approximate the surface area S_i of this region of the surface with the area T_i of the corresponding portion of the tangent plane.

This portion of the tangent plane is a parallelogram, defined by sides $\vec{u}$ and $\vec{v}$, as shown. One of the applications of the cross product from Section 10.4 is that the area of this parallelogram is $\|\vec{u} \times \vec{v}\|$. Once we can determine $\vec{u}$ and $\vec{v}$, we can determine the area.

$\vec{u}$ is tangent to the surface in the direction of x, therefore, from Section 12.7, $\vec{u}$ is parallel to $\langle 1, 0, f_x(x_i, y_i)\rangle$. The x-displacement of $\vec{u}$ is Δx_i, so we know that $\vec{u} = \Delta x_i \langle 1, 0, f_x(x_i, y_i)\rangle$. Similar logic shows that $\vec{v} = \Delta y_i \langle 0, 1, f_y(x_i, y_i)\rangle$. Thus:

$$\begin{aligned}\text{surface area } S_i &\approx \text{area of } T_i \\ &= \|\vec{u} \times \vec{v}\| \\ &= \big|\big|\Delta x_i \langle 1, 0, f_x(x_i, y_i)\rangle \times \Delta y_i \langle 0, 1, f_y(x_i, y_i)\rangle\big|\big| \\ &= \sqrt{1 + f_x(x_i, y_i)^2 + f_y(x_i, y_i)^2}\, \Delta x_i\, \Delta y_i.\end{aligned}$$

Note that $\Delta x_i\, \Delta y_i = \Delta A_i$, the area of the i^{th} subregion.

Summing up all n of the approximations to the surface area gives

$$\text{surface area over } R \approx \sum_{i=1}^{n} \sqrt{1 + f_x(x_i, y_i)^2 + f_y(x_i, y_i)^2}\, \Delta A_i. \tag{13.5.1}$$

Once again take a limit as all of the Δx_i and Δy_i shrink to 0; this leads to a double integral.

Definition 105 Surface Area

Let $z = f(x, y)$ where f_x and f_y are continuous over a closed, bounded region R. The surface area S over R is

$$\begin{aligned} S &= \iint_R \, dS \\ &= \iint_R \sqrt{1 + f_x(x, y)^2 + f_y(x, y)^2} \, dA. \end{aligned}$$

Note: as done before, we think of "$\iint_R \, dS$" as meaning "sum up lots of little surface areas over R."

The concept of surface area is *defined* here, for while we already have a notion of the area of a region in the *plane*, we did not yet have a solid grasp of what "the area of a surface in *space*" means.

We test this definition by using it to compute surface areas of known surfaces. We start with a triangle.

Example 13.5.1: Finding the surface area of a plane over a triangle

Let $f(x, y) = 4 - x - 2y$, and let R be the region in the plane bounded by $x = 0$, $y = 0$ and $y = 2 - x/2$, as shown in Figure 13.5.2. Find the surface area of f over R.

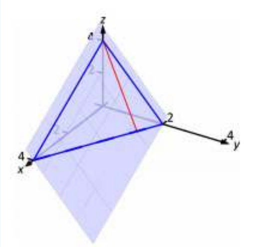


Figure 13.5.2: Finding the area of a triangle in space in Example 13.5.1.

Solution

We follow Definition 105. We start by noting that $f_x(x, y) = -1$ and $f_y(x, y) = -2$. To define R, we use bounds $0 \le y \le 2 - x/2$ and $0 \le x \le 4$. Therefore

$$\begin{aligned} S &= \iint_R \, dS \\ &= \int_0^4 \int_0^{2-x/2} \sqrt{1 + (-1)^2 + (-2)^2} \, dy \, dx \\ &= \int_0^4 \sqrt{6} \left(2 - \frac{x}{2}\right) dx \\ &= 4\sqrt{6}. \end{aligned}$$

Because the surface is a triangle, we can figure out the area using geometry. Considering the base of the triangle to be the side in the xy-plane, we find the length of the base to be $\sqrt{20}$. We can find the height using our knowledge of vectors: let $\vec{u}$ be the side in the x-z plane and let $\vec{v}$ be the side in the xy-plane. The height is then $\|\vec{u} - \text{proj}_{\vec{v}}\vec{u}\| = 4\sqrt{6/5}$. Geometry states that the area is thus

$$\frac{1}{2} \cdot 4\sqrt{6/5} \cdot \sqrt{20} = 4\sqrt{6}. \tag{13.5.2}$$

We affirm the validity of our formula.

It is "common knowledge" that the surface area of a sphere of radius r is $4\pi r^2$. We confirm this in the following example, which involves using our formula with polar coordinates.

Example 13.5.2: The surface area of a sphere.

Find the surface area of the sphere with radius a centered at the origin, whose top hemisphere has equation $f(x, y) = \sqrt{a^2 - x^2 - y^2}$.

Solution

We start by computing partial derivatives and find

$$f_x(x,y) = \frac{-x}{\sqrt{a^2-x^2-y^2}} \quad \text{and} \quad f_y(x,y) = \frac{-y}{\sqrt{a^2-x^2-y^2}}. \tag{13.5.3}$$

As our function f only defines the top upper hemisphere of the sphere, we double our surface area result to get the total area:

$$\begin{aligned} S &= 2\iint_R \sqrt{1+f_x(x,y)^2+f_y(x,y)^2}\, dA \\ &= 2\iint_R \sqrt{1+\frac{x^2+y^2}{a^2-x^2-y^2}}\, dA. \end{aligned}$$

The region R that we are integrating over is the circle, centered at the origin, with radius a: $x^2+y^2=a^2$. Because of this region, we are likely to have greater success with our integration by converting to polar coordinates. Using the substitutions $x = r\cos\theta$, $y = r\sin\theta$, $dA = r\ dr\ d\theta$ and bounds $0 \le \theta \le 2\pi$ and $0 \le r \le a$, we have:

$$\begin{aligned} S &= 2\int_0^{2\pi}\int_0^a \sqrt{1+\frac{r^2\cos^2\theta+r^2\sin^2\theta}{a^2-r^2\cos^2\theta-r^2\sin^2\theta}}\ r\ dr\ d\theta \\ &= 2\int_0^{2\pi}\int_0^a r\sqrt{1+\frac{r^2}{a^2-r^2}}\ dr\ d\theta \\ &= 2\int_0^{2\pi}\int_0^a r\sqrt{\frac{a^2}{a^2-r^2}}\ dr\ d\theta. \end{aligned} \tag{13.5.4}$$

Apply substitution $u = a^2 - r^2$ and integrate the inner integral, giving

$$\begin{aligned} &= 2\int_0^{2\pi} a^2\ d\theta \\ &= 4\pi a^2 \end{aligned} .$$

Our work confirms our previous formula.

Note: The inner integral in Equation 13.5.4 is an improper integral, as the integrand of $\int_0^a r\sqrt{\frac{a^2}{a^2-r^2}}\ dr$ is not defined at $r=a$. To properly evaluate this integral, one must use the techniques of Section 6.8.

The reason this need arises is that the function $f(x,y)=\sqrt{a^2-x^2-y^2}$ fails the requirements of Definition 105, as f_x and f_y are not continuous on the boundary of the circle $x^2+y^2=a^2$.

The computation of the surface area is still valid. The definition makes stronger requirements than necessary in part to avoid the use of improper integration, as when f_x and/or f_y are not continuous, the resulting improper integral may not converge. Since the improper integral does converge in this example, the surface area is accurately computed.

Example 13.5.3: Finding the surface area of a cone

The general formula for a right cone with height h and base radius a is $f(x,y) = h - \frac{h}{a}\sqrt{x^2+y^2}$, shown in Figure 13.34. Find the surface area of this cone.

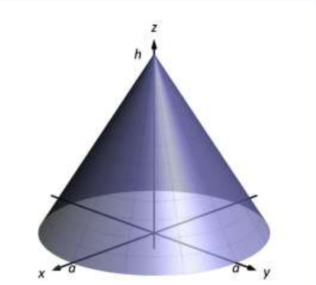


Figure 13.5.3: Finding the surface area of a cone in Example 13.5.3.

Solution

We begin by computing partial derivatives.

$$f_x(x,y) = -\frac{xh}{a\sqrt{x^2+y^2}} \quad \text{and} \quad -\frac{yh}{a\sqrt{x^2+y^2}}.$$

Since we are integrating over the circle $x^2+y^2=a^2$, we again use polar coordinates. Using the standard substitutions, our integrand becomes

$$\sqrt{1+\left(\frac{hr\cos\theta}{a\sqrt{r^2}}\right)^2+\left(\frac{hr\sin\theta}{a\sqrt{r^2}}\right)^2}.$$

This may look intimidating at first, but there are lots of simple simplifications to be done. It amazingly reduces to just

$$\sqrt{1+\frac{h^2}{a^2}} = \frac{1}{a}\sqrt{a^2+h^2}.$$

Our polar bounds are $0 \le \theta \le 2\pi$ and $0 \le r \le a$. Thus

$$\begin{aligned} S &= \int_0^{2\pi}\int_0^a r\frac{1}{a}\sqrt{a^2+h^2}\,dr\,d\theta \\ &= \int_0^{2\pi}\left(\frac{1}{2}r^2\frac{1}{a}\sqrt{a^2+h^2}\right)\Big|_0^a\,d\theta \\ &= \int_0^{2\pi}\frac{1}{2}a\sqrt{a^2+h^2}\,d\theta \\ &= \pi a\sqrt{a^2+h^2}. \end{aligned}$$

This matches the formula found in the back of this text.

Note: Note that once again f_x and f_y are not continuous on the domain of f, as both are undefined at $(0,0)$. (A similar problem occurred in the previous example.) Once again the resulting improper integral converges and the computation of the surface area is valid.

Example $13.5.4$: Finding surface area over a region

Find the area of the surface $f(x,y)=x^2-3y+3$ over the region R bounded by $-x \le y \le x$, $0 \le x \le 4$, as pictured in Figure $13.5.4$.

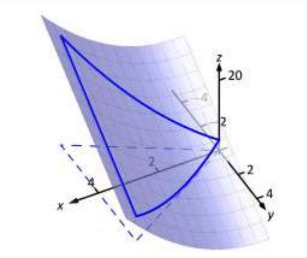


Figure $13.5.4$: Graphing the surface in Example $13.5.4$.

Solution

It is straightforward to compute $f_x(x,y)=2x$ and $f_y(x,y)=-3$. Thus the surface area is described by the double integral

$$\iint_R \sqrt{1+(2x)^2+(-3)^2}\,dA = \iint_R \sqrt{10+4x^2}\,dA.$$

As with integrals describing arc length, double integrals describing surface area are in general hard to evaluate directly because of the square-root. This particular integral can be easily evaluated, though, with judicious choice of our order of integration.

Integrating with order $dx\,dy$ requires us to evaluate $\int\sqrt{10+4x^2}dx$. This can be done, though it involves Integration By Parts and $\sinh^{-1}x$. Integrating with order $dy\,dx$ has as its first integral $\int\sqrt{10+4x^2}\,dy$, which is easy to evaluate: it is simply $y\sqrt{10+4x^2}+C$. So we proceed with the order $dy\,dx$; the bounds are already given in the statement of the problem.

$$\begin{aligned}\iint_R \sqrt{10+4x^2}\,dA &= \int_0^4 \int_{-x}^{x} \sqrt{10+4x^2}\,dy\,dx \\ &= \int_0^4 \left(y\sqrt{10+4x^2}\right)\Big|_{-x}^{x}\,dx \\ &= \int_0^4 \left(2x\sqrt{10+4x^2}\right)dx.\end{aligned}$$

Apply substitution with $u = 10+4x^2$:

$$\begin{aligned}&= \left(\frac{1}{6}\left(10+4x^2\right)^{3/2}\right)\Big|_0^4 \\ &= \frac{1}{3}\left(37\sqrt{74}-5\sqrt{10}\right) \approx 100.825 \text{ units}^2.\end{aligned}$$

So while the region R over which we integrate has an area of 16 units^2, the surface has a much greater area as its z-values change dramatically over R.

In practice, technology helps greatly in the evaluation of such integrals. High powered computer algebra systems can compute integrals that are difficult, or at least time consuming, by hand, and can at the least produce very accurate approximations with numerical methods. In general, just knowing *how* to set up the proper integrals brings one very close to being able to compute the needed value. Most of the work is actually done in just describing the region R in terms of polar or rectangular coordinates. Once this is done, technology can usually provide a good answer.

We have learned how to integrate integrals; that is, we have learned to evaluate double integrals. In the next section, we learn how to integrate double integrals -- that is, we learn to evaluate *triple integrals*, along with learning some uses for this operation.

This page titled 13.5: Surface Area is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

13.6: Volume Between Surfaces and Triple Integration

We learned in Section 13.2 how to compute the signed volume V under a surface $z=f(x,y)$ over a region R: $V=\iint_R f(x,y)dA$. It follows naturally that if $f(x,y)\geq g(x,y)$ on R, then the **volume between** $f(x,y)$ and $g(x,y)$ on R is

$$V=\iint_R f(x,y)dA-\iint_R g(x,y)dA \tag{13.6.1}$$

$$=\iint_R \big(f(x,y)-g(x,y)\big)dA. \tag{13.6.2}$$

theorem 124: Volume Between Surfaces

Let f and g be continuous functions on a closed, bounded region R, where $f(x,y)\geq g(x,y)$ for all (x,y) in R. The volume V between f and g over R is

$$V=\iint_R \big(f(x,y)-g(x,y)\big)dA. \tag{13.6.3}$$

Example 13.6.1: Finding volume between surfaces

Find the volume of the space region bounded by the planes $z=3x+y-4$ and $z=8-3x-2y$ in the 1^{st} octant. In Figure 13.36(a) the planes are drawn; in (b), only the defined region is given.

Solution

We need to determine the region R over which we will integrate. To do so, we need to determine where the planes intersect. They have common z-values when $3x+y-4=8-3x-2y$. Applying a little algebra, we have:

$$\begin{aligned}3x+y-4&=8-3x-2y\\6x+3y&=12\\2x+y&=4\end{aligned}$$

The planes intersect along the line $2x+y=4$. Therefore the region R is bounded by $x=0$, $y=0$, and $y=4-2x$; we can convert these bounds to integration bounds of $0\leq x\leq 2$, $0\leq y\leq 4-2x$. Thus

$$\begin{aligned}V&=\iint_R \big(8-3x-2y-(3x+y-4)\big)dA\\&=\int_0^2\int_0^{4-2x}\big(12-6x-3y\big)dy\,dx\\&=16\text{u}^3.\end{aligned}$$

The volume between the surfaces is 16 cubic units.

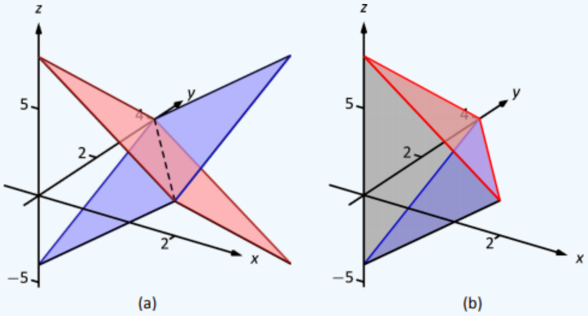


Figure 13.36: Finding the volume between the planes given in Example 13.6.1.

In the preceding example, we found the volume by evaluating the integral

$$\int_0^2\int_0^{4-2x}\big(8-3x-2y-(3x+y-4)\big)dy\,dx. \tag{13.6.4}$$

Note how we can rewrite the integrand as an integral, much as we did in Section 13.1:

$$\int^{8-3x-2y}$$

$$8-3x-2y-(3x+y-4)=\int_{3x+y-4} dz. \tag{13.6.5}$$

Thus we can rewrite the double integral that finds volume as

$$\int_0^2\int_0^{4-2x}\big(8-3x-2y-(3x+y-4)\big)dy\,dx=\int_0^2\int_0^{4-2x}\left(\int_{3x+y-4}^{8-3x-2y} dz\right)dy\,dx. \tag{13.6.6}$$

This no longer looks like a "double integral," but more like a "triple integral." Just as our first introduction to double integrals was in the context of finding the area of a plane region, our introduction into triple integrals will be in the context of finding the volume of a space region.

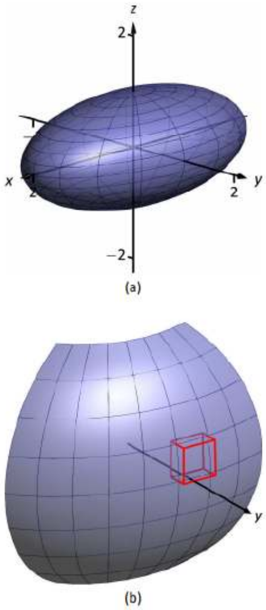


Figure 13.37: Approximating the volume of a region D in space.

To formally find the volume of a closed, bounded region D in space, such as the one shown in Figure 13.37(a), we start with an approximation. Break D into n rectangular solids; the solids near the boundary of D may possibly not include portions of D and/or include extra space. In Figure 13.37(b), we zoom in on a portion of the boundary of D to show a rectangular solid that contains space not in D; as this is an approximation of the volume, this is acceptable and this error will be reduced as we shrink the size of our solids.

The volume ΔV_i of the i^{th} solid D_i is $\Delta V_i=\Delta x_i\Delta y_i\Delta z_i$, where Δx_i, Δy_i and Δz_i give the dimensions of the rectangular solid in the x, y and z directions, respectively. By summing up the volumes of all n solids, we get an approximation of the volume V of D:
$$V \approx \sum_{i=1}^n \Delta V_i = \sum_{i=1}^n \Delta x_i\Delta y_i\Delta z_i.\]

Let $||\Delta D||$ represent the length of the longest diagonal of rectangular solids in the subdivision of D. As $||\Delta D||\to 0$, the volume of each solid goes to 0, as do each of Δx_i, Δy_i and Δz_i, for all i. Our calculus experience tells us that taking a limit as $||\Delta D||\to 0$ turns our approximation of V into an exact calculation of V. Before we state this result in a theorem, we use a definition to define some terms.

Definition 106: Triple Integrals, Iterated Integration (Part I)

Let D be a closed, bounded region in space. Let a and b be real numbers, let $g_1(x)$ and $g_2(x)$ be continuous functions of x, and let $f_1(x,y)$ and $f_2(x,y)$ be continuous functions of x and y.

1. The volume V of D is denoted by a **triple integral**,

$$V=\iiint_D dV. \tag{13.6.7}$$

2. The iterated integral $\int_a^b\int_{g_1(x)}^{g_2(x)}\int_{f_1(x,y)}^{f_2(x,y)} dz\,dy\,dx$ is evaluated as

$$\int^b\int^{g_2(x)}\int^{f_2(x,y)} dz\,dy\,dx=\int^b\int^{g_2(x)}\left(\int^{f_2(x,y)} dz\right)dy\,dx \tag{13.6.8}$$

$$\int_a \int_{g_1(x)} \int_{f_1(x,y)} dz\,dy\,dx = \int_a \int_{g_1(x)} \left(\int_{f_1(x,y)} dz\right) dy\,dx.$$

Evaluating the above iterated integral is **triple integration**.

Our informal understanding of the notation $\iiint_D dV$ is "sum up lots of little volumes over D," analogous to our understanding of $\iint_R dA$ and $\iint_R \ dm$.

We now state the major theorem of this section.

theorem 125 Triple Integration (Part I)

Let D be a closed, bounded region in space and let ΔD be any subdivision of D into n rectangular solids, where the i^{th} subregion D_i has dimensions $\Delta x_i \times \Delta y_i \times \Delta z_i$ and volume ΔV_i.

1. The volume V of D is

$$V = \iiint_D dV = \lim_{||\Delta D|| \to 0} \sum_{i=1}^{n} \Delta V_i = \lim_{||\Delta D|| \to 0} \sum_{i=1}^{n} \Delta x_i \Delta y_i \Delta z_i. \tag{13.6.9}$$

2. If D is defined as the region bounded by the planes $x = a$ and $x = b$, the cylinders $y = g_{(}x)$ and $y = g_2(x)$, and the surfaces $z = f_1(x,y)$ and $z = f_2(x,y)$, where $a < b$, $g_1(x) \leq g_2(x)$ and $f_1(x,y) \leq f_2(x,y)$ on D, then

$$\iiint_D dV = \int_a^b \int_{g_1(x)}^{g_2(x)} \int_{f_1(x,y)}^{f_2(x,y)} dz\,dy\,dx. \tag{13.6.10}$$

3. V can be determined using iterated integration with other orders of integration (there are 6 total), as long as D is defined by the region enclosed by a pair of planes, a pair of cylinders, and a pair of surfaces.

We evaluated the area of a plane region R by iterated integration, where the bounds were "from curve to curve, then from point to point." Theorem 125 allows us to find the volume of a space region with an iterated integral with bounds "from surface to surface, then from curve to curve, then from point to point." In the iterated integral

$$\int_a^b \int_{g_1(x)}^{g_2(x)} \int_{f_1(x,y)}^{f_2(x,y)} dz\,dy\,dx, \tag{13.6.11}$$

the bounds $a \leq x \leq b$ and $g_1(x) \leq y \leq g_2(x)$ define a region R in the x-y plane over which the region D exists in space. However, these bounds are also defining surfaces in space; $x = a$ is a plane and $y = g_1(x)$ is a cylinder. The combination of these 6 surfaces enclose, and define, D.

Examples will help us understand triple integration, including integrating with various orders of integration.

Example 13.6.2: Finding the volume of a space region with triple integration

Find the volume of the space region in the 1^{st} octant bounded by the plane $z = 2 - y/3 - 2x/3$, shown in Figure 13.38(a), using the order of integration $dz\,dy\,dx$. Set up the triple integrals that give the volume in the other 5 orders of integration.

Solution

Starting with the order of integration $dz\,dy\,dx$, we need to first find bounds on z. The region D is bounded below by the plane $z = 0$ (because we are restricted to the first octant) and above by $z = 2 - y/3 - 2x/3$; $0 \leq z \leq 2 - y/3 - 2x/3$.

To find the bounds on y and x, we "collapse" the region onto the x-y plane, giving the triangle shown in Figure 13.38(b). (We know the equation of the line $y = 6 - 2x$ in two ways. First, by setting $z = 0$, we have $0 = 2 - y/3 - 2x/3 \Rightarrow y = 6 - 2x$. Secondly, we know this is going to be a straight line between the points $(3,0)$ and $(0,6)$ in the x-y plane.)

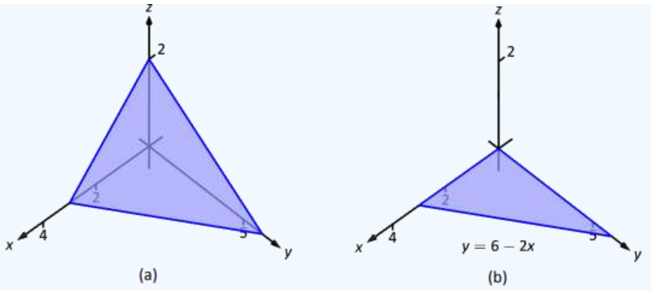


Figure 13.38: The region D used in Example 13.6.2 in (a); in (b), the region found by collapsing D onto the x-y plane.

We define that region R, in the integration order of $dy\,dx$, with bounds $0 \le y \le 6 - 2x$ and $0 \le x \le 3$. Thus the volume V of the region D is:

$$\begin{aligned} V &= \iiint_D dV \\ &= \int_0^3 \int_0^{6-2x} \int_0^{2-\frac{1}{3}y-\frac{2}{3}x} dzdydz \\ &= \int_0^3 \int_0^{6-2x} \left(\int_0^{2-\frac{1}{3}y-\frac{2}{3}x} dz \right) dydz \\ &= \int_0^3 \int_0^{6-2x} z\Big|_0^{2-\frac{1}{3}y-\frac{2}{3}x} dydz \\ &= \int_0^3 \int_0^{6-2x} \left(2 - \frac{1}{3}y - \frac{2}{3}x \right) dydz. \end{aligned}$$

From this step on, we are evaluating a double integral as done many times before. We skip these steps and give the final volume,

$$= 6\mathrm{u}^3. \tag{13.6.12}$$

The order $dz\,dx\,dy$:

Now consider the volume using the order of integration $dz\,dx\,dy$. The bounds on z are the same as before, $0 \le z \le 2 - y/3 - 2x/3$. Collapsing the space region on the x-y plane as shown in Figure 13.38(b), we now describe this triangle with the order of integration $dxdy$. This gives bounds $0 \le x \le 3 - y/2$ and $0 \le y \le 6$. Thus the volume is given by the triple integral

$$V = \int_0^6 \int_0^{3-\frac{1}{2}y} \int_0^{2-\frac{1}{3}y-\frac{2}{3}x} dz\,dx\,dy. \tag{13.6.13}$$

The order $dx\,dy\,dz$:

Following our "surface to surface..." strategy, we need to determine the x-*surfaces* that bound our space region. To do so, approach the region "from behind," in the direction of increasing x. The first surface we hit as we enter the region is the y-z plane, defined by $x = 0$. We come out of the region at the plane $z = 2 - y/3 - 2x/3$; solving for x, we have $x = 3 - y/2 - 3z/2$. Thus the bounds on x are: $0 \le x \le 3 - y/2 - 3z/2$.

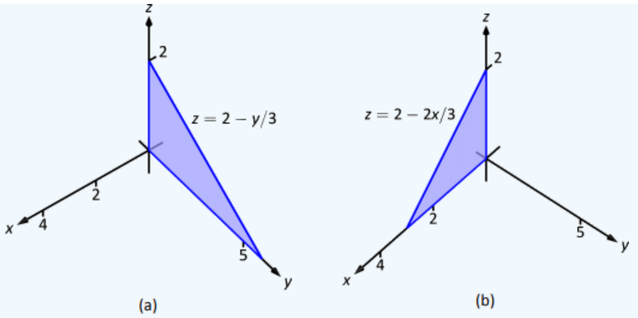


Figure 13.39: The region D in Example 13.6.2 is collapsed onto the y-z plane in (a); in (b), the region is collapsed onto the x-z plane.

Now collapse the space region onto the y-z plane, as shown in Figure 13.39(a). (Again, we find the equation of the line $z = 2 - y/3$ by setting $x = 0$ in the equation $x = 3 - y/2 - 3z/2$.) We need to find bounds on this region with the order $dydz$. The *curves* that bound y are $y = 0$ and $y = 6 - 3z$; the *points* that bound z are 0 and 2. Thus the triple integral giving volume is:

$$\begin{array}{c} 0 \le x \le 3 - y/2 - 3z/2 \\ 0 \le y \le 6 - 3z \\ 0 \le z \le 2 \end{array} \quad \Rightarrow \quad \int_0^2 \int_0^{6-3z} \int_0^{3-y/2-3z/2} dx\, dy\, dz. \tag{13.6.14}$$

The order $dx\,dz\,dy$:

The x-bounds are the same as the order above. We now consider the triangle in Figure 13.39(a) and describe it with the order $dzdy$: $0 \le z \le 2 - y/3$ and $0 \le y \le 6$. Thus the volume is given by:
$$

Missing \end{array}

&
\Rightarrow \quad \int_0^6\int_0^{2-y/3}\int_0^{3-y/2-3z/2} dx \, dz \, dy.
\end{array}
\]

The order $dydzdx$:

We now need to determine the y-surfaces that determine our region. Approaching the space region from "behind" and moving in the direction of increasing y, we first enter the region at $y = 0$, and exit along the plane $z = 2 - y/3 - 2x/3$. Solving for y, this plane has equation $y = 6 - 2x - 3z$. Thus y has bounds $0 \le y \le 6 - 2x - 3z$.

Now collapse the region onto the x-z plane, as shown in Figure 13.39(b). The curves bounding this triangle are $z = 0$ and $z = 2 - 2x/3$; x is bounded by the points $x = 0$ to $x = 3$. Thus the triple integral giving volume is:
$$

Missing \end{array}

&
\Rightarrow \quad \int_0^3\int_0^{2-2x/3}\int_0^{6-2x-3z} dy dz dx.
\end{array}
\]

The order $dy\,dx\,dz$:

The y-bounds are the same as in the order above. We now determine the bounds of the triangle in Figure 13.39(b) using the order $dy\,dx\,dz$. x is bounded by $x = 0$ and $x = 3 - 3z/2$; z is bounded between $z = 0$ and $z = 2$. This leads to the triple integral:

$$\begin{array}{c} 0 \le y \le 6 - 2x - 3z \\ 0 \le x \le 3 - 3z/2 \\ 0 \le z \le 2 \end{array} \quad \Rightarrow \quad \int_0^2 \int_0^{3-3z/2} \int_0^{6-2x-3z} dy\, dx\, dz. \tag{13.6.15}$$

This problem was long, but hopefully useful, demonstrating how to determine bounds with every order of integration to describe the region D. In practice, we only need 1, but being able to do them all gives us flexibility to choose the order that suits us best.

In the previous example, we collapsed the surface into the x-y, x-z, and y-z planes as we determined the "curve to curve, point to point" bounds of integration. Since the surface was a triangular portion of a plane, this collapsing, or *projecting*, was simple: the *projection* of a straight line in space onto a coordinate plane is a line.

The following example shows us how to do this when dealing with more complicated surfaces and curves.

Example 13.6.3: Finding the projection of a curve in space onto the coordinate planes

Consider the surfaces $z=3-x^2-y^2$ and $z=2y$, as shown in Figure 13.40(a). The curve of their intersection is shown, along with the projection of this curve into the coordinate planes, shown dashed. Find the equations of the projections into the coordinate planes.

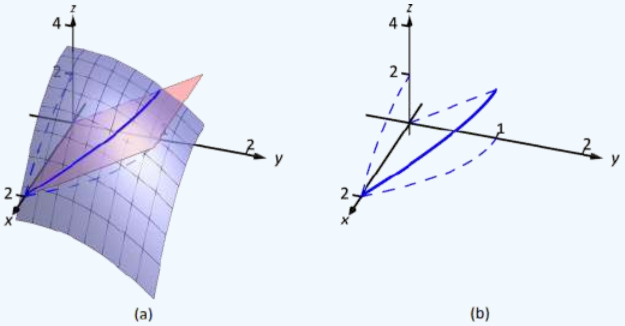


Figure 13.40: Finding the projections of the curve of intersection in Example 13.6.3.

Solution

The two surfaces are $z=3-x^2-y^2$ and $z=2y$. To find where they intersect, it is natural to set them equal to each other: $3-x^2-y^2=2y$. This is an implicit function of x and y that gives all points (x,y) in the x-y plane where the z values of the two surfaces are equal.

We can rewrite this implicit function by completing the square:

$$3-x^2-y^2=2y \quad \Rightarrow \quad y^2+2y+x^2=3 \quad \Rightarrow \quad (y+1)^2+x^2=4. \tag{13.6.16}$$

Thus in the x-y plane the projection of the intersection is a circle with radius 2, centered at $(0,-1)$.

To project onto the x-z plane, we do a similar procedure: find the x and z values where the y values on the surface are the same. We start by solving the equation of each surface for y. In this particular case, it works well to actually solve for y^2:

$z=3-x^2-y^2 \quad \Rightarrow \quad y^2=3-x^2-z$
$z=2y \quad \Rightarrow \quad y^2=z^2/4.$

Thus we have (after again completing the square):

$$3-x^2-z=z^2/4 \quad \Rightarrow \quad \frac{(z+2)^2}{16}+\frac{x^2}{4}=1, \tag{13.6.17}$$

and ellipse centered at $(0,-2)$ in the x-z plane with a major axis of length 8 and a minor axis of length 4.

Finally, to project the curve of intersection into the y-z plane, we solve equation for x. Since $z=2y$ is a cylinder that lacks the variable x, it becomes our equation of the projection in the y-z plane.

All three projections are shown in Figure 13.40(b).

Example $13.6.4$: Finding the volume of a space region with triple integration

Set up the triple integrals that find the volume of the space region D bounded by the surfaces $x^2+y^2=1$, $z=0$ and $z=-y$, as shown in Figure 13.41(a), with the orders of integration $dz\,dy\,dx$, $dy\,dx\,dz$ and $dx\,dz\,dy$.

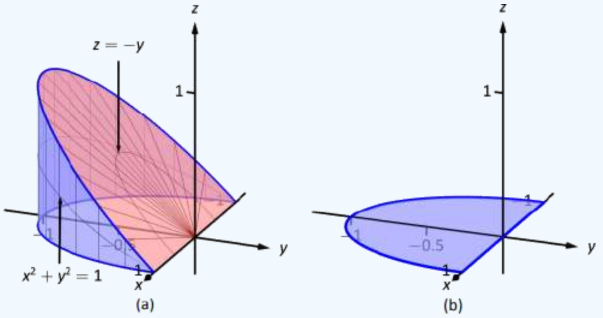


Figure 13.41: The region D in Example 13.6.4 is shown in (a); in (b), it is collapsed onto the x-y plane.

Solution

The order $dz\,dy\,dx$:

The region D is bounded below by the plane $z=0$ and above by the plane $z=-y$. The cylinder $x^2+y^2=1$ does not offer any bounds in the z-direction, as that surface is parallel to the z-axis. Thus $0 \leq z \leq -y$.

Collapsing the region into the x-y plane, we get part of the circle with equation $x^2+y^2=1$ as shown in Figure 13.41(b). As a function of x, this half circle has equation $y=-\sqrt{1-x^2}$. Thus y is bounded below by $-\sqrt{1-x^2}$ and above by $y=0$: $-\sqrt{1-x^2} \leq y \leq 0$. The x bounds of the half circle are $-1 \leq x \leq 1$. All together, the bounds of integration and triple integral are as follows:

$$\begin{array}{c} 0 \leq z \leq -y \\ -\sqrt{1-x^2} \leq y \leq 0 \\ -1 \leq x \leq 1 \end{array} \quad \Rightarrow \quad \int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{0}\int_{0}^{-y} dz\,dy\,dx. \tag{13.6.18}$$

We evaluate this triple integral:

$$\begin{aligned} \int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{0}\int_{0}^{-y} dz\,dy\,dx &= \int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{0}\left(-y\right)dy\,dx \\ &= \int_{-1}^{1}\left(-\frac{1}{2}y^2\right)\Big|_{-\sqrt{1-x^2}}^{0}dx \\ &= \int_{-1}^{1}\frac{1}{2}\left(1-x^2\right)dx \\ &= \left(\frac{1}{2}\left(x-\frac{1}{3}x^3\right)\right)\Big|_{-1}^{1} \\ &= \frac{2}{3}\text{units}^3. \end{aligned}$$

With the order $dy\,dx\,dz$:

The region is bounded "below" in the y-direction by the surface $x^2+y^2=1 \Rightarrow y=-\sqrt{1-x^2}$ and "above" by the surface $y=-z$. Thus the y bounds are $-\sqrt{1-x^2} \leq y \leq -z$.

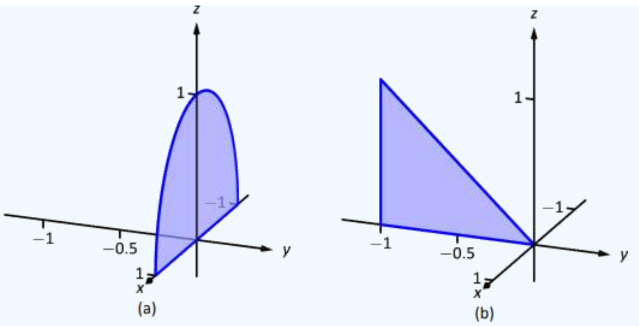


Figure 13.42: The region D in Example 13.6.4 is shown collapsed onto the x-z plane in (a); in (b), it is collapsed onto the y-z plane.

Collapsing the region onto the x-z plane gives the region shown in Figure 13.42(a); this half circle has equation $x^2+z^2=1$. (We find this curve by solving each surface for y^2, then setting them equal to each other. We have $y^2=1-x^2$ and $y=-z \Rightarrow y^2=z^2$. Thus $x^2+z^2=1$.) It is bounded below by $x=-\sqrt{1-z^2}$ and above by $x=\sqrt{1-z^2}$, where z is bounded by $0 \le z \le 1$. All together, we have:

$$\begin{array}{c} -\sqrt{1-x^2} \le y \le -z \\ -\sqrt{1-z^2} \le x \le \sqrt{1-z^2} \\ 0 \le z \le 1 \end{array} \quad \Rightarrow \quad \int_0^1 \int_{-\sqrt{1-z^2}}^{\sqrt{1-z^2}} \int_{-\sqrt{1-x^2}}^{-z} dy\,dx\,dz. \tag{13.6.19}$$

With the order $dx\,dz\,dy$:

D is bounded below by the surface $x=-\sqrt{1-y^2}$ and above by $\sqrt{1-y^2}$. We then collapse the region onto the y-z plane and get the triangle shown in Figure 13.42}(b). (The hypotenuse is the line $z=-y$, just as the plane.) Thus z is bounded by $0 \le z \le -y$ and y is bounded by $-1 \le y \le 0$. This gives:

$$\begin{array}{c} -\sqrt{1-y^2} \le x \le \sqrt{1-y^2} \\ 0 \le z \le -y \\ -1 \le y \le 0 \end{array} \quad \Rightarrow \quad \int_{-1}^0 \int_0^{-y} \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} dx\,dz\,dy. \tag{13.6.20}$$

The following theorem states two things that should make "common sense" to us. First, using the triple integral to find volume of a region D should always return a positive number; we are computing *volume* here, not *signed volume*. Secondly, to compute the volume of a "complicated" region, we could break it up into subregions and compute the volumes of each subregion separately, summing them later to find the total volume.

THEOREM 126: Properties of Triple Integrals

Let D be a closed, bounded region in space, and let D_1 and D_2 be non-overlapping regions such that $D=D_1 \bigcup D_2$.

1. $\iiint_D dV \ge 0$
2. $\iiint_D dV = \iiint_{D_1} dV + \iiint_{D_2} dV$.

We use this latter property in the next example.

Example 13.6.5: Finding the volume of a space region with triple integration

Find the volume of the space region D bounded by the coordinate planes, $z=1-x/2$ and $z=1-y/4$, as shown in Figure 13.43(a). Set up the triple integrals that find the volume of D in all 6 orders of integration.

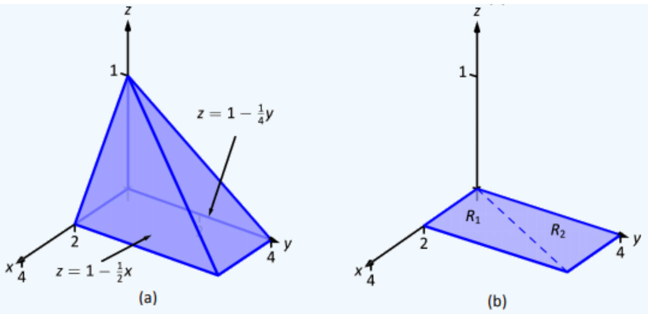


Figure 13.43: The region D in Example 13.6.5 is shown in (a); in (b), it is collapsed onto the x-y plane.

Solution
Following the bounds--determining strategy of "surface to surface, curve to curve, and point to point," we can see that the most difficult orders of integration are the two in which we integrate with respect to z first, for there are two "upper" surfaces that bound D in the z-direction. So we start by noting that we have

$$0 \le z \le 1 - \frac{1}{2}x \quad \text{and} \quad 0 \le z \le 1 - \frac{1}{4}y. \tag{13.6.21}$$

We now collapse the region D onto the x-y axis, as shown in Figure 13.43(b). The boundary of D, the line from $(0, 0, 1)$ to $(2, 4, 0)$, is shown in part (b) of the figure as a dashed line; it has equation $y = 2x$. (We can recognize this in two ways: one, in collapsing the line from $(0, 0, 1)$ to $(2, 4, 0)$ onto the x-y plane, we simply ignore the z-values, meaning the line now goes from $(0, 0)$ to $(2, 4)$. Secondly, the two surfaces meet where $z = 1 - x/2$ is equal to $z = 1 - y/4$: thus $1 - x/2 = 1 - y/4 \Rightarrow y = 2x$.)

We use the second property of Theorem 126 to state that

$$\iiint_D dV = \iiint_{D_1} dV + \iiint_{D_2} dV, \tag{13.6.22}$$

where D_1 and D_2 are the space regions above the plane regions R_1 and R_2, respectively. Thus we can say

$$\iiint_D dV = \iint_{R_1} \left(\int_0^{1-x/2} dz \right) dA + \iint_{R_2} \left(\int_0^{1-y/4} dz \right) dA. \tag{13.6.23}$$

All that is left is to determine bounds of R_1 and R_2, depending on whether we are integrating with order $dx dy$ or $dy\, dx$. We give the final integrals here, leaving it to the reader to confirm these results.

$dz\, dy\, dx$:
$$

Missing \end{array}

& &

$$\begin{array}{c} 0 \le z \le 1 - y/4 \\ 2x \le y \le 4 \\ 0 \le x \le 2 \end{array} \tag{13.6.24}$$

\
\
\iiint_D dV &=& \int_0^2\int_0^{2x}\int_0^{1-x/2} dz \, dy \, dx &+& \int_0^2\int_{2x}^4\int_0^{1-y/4} dz \, dy \, dx
\end{array}
\]

$dz\, dx\, dy$:
$$

Missing \end{array}

& &

$$\begin{array}{c} 0 \le z \le 1 - y/4 \\ 0 \le x \le y/2 \\ 0 \le y \le 4 \end{array} \tag{13.6.25}$$

\
\
\iiint_D dV &=& \int_0^4\int_{y/2}^{2}\int_0^{1-x/2} dz \, dx \, dy &+& \int_0^4\int_{0}^{y/2}\int_0^{1-y/4} dz \, dx \, dy
\end{array}
\]

The remaining four orders of integration do not require a sum of triple integrals. In Figure 13.44 we show D collapsed onto the other two coordinate planes. Using these graphs, we give the final orders of integration here, again leaving it to the reader to confirm these results.

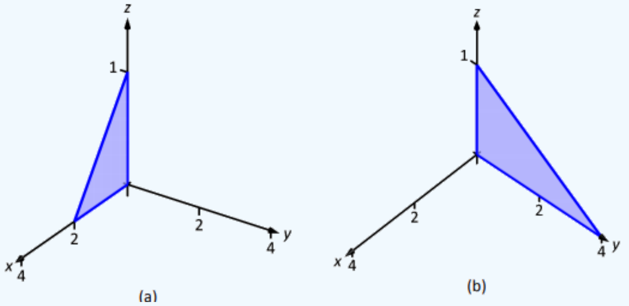


Figure 13.44: The region D in Example 13.6.5 is shown collapsed onto the x-z plane in (a); in (b), it is collapsed onto the y-z plane.

$dy\,dx\,dz$:

$$\begin{array}{c} 0 \le y \le 4 - 4z \\ 0 \le x \le 2 - 2z \\ 0 \le z \le 1 \end{array} \quad \Rightarrow \int_0^1\int_0^{2-2z}\int_0^{4-4z} dy\,dx\,dz \tag{13.6.26}$$

$dydzdx$:
$$

Missing \end{array}

&
\Rightarrow \int_0^2\int_{0}^{1-x/2}\int_0^{4-4z} dy \, dx \, dz
\end{array}
\]

$dx\,dy\,dz$:
$$

Missing \end{array}

&
\Rightarrow \int_0^1\int_{0}^{4-4z}\int_0^{2-2z} dx \, dy \, dz
\end{array}
\]

$dx\,dz\,dy$:
$$

Missing \end{array}

&
\Rightarrow \int_0^4\int_{0}^{1-y/4}\int_0^{2-2z} dx \, dz \, dy
\end{array}
\]

We give one more example of finding the volume of a space region.

Example 13.6.6: Finding the volume of a space region

Set up a triple integral that gives the volume of the space region D bounded by $z = 2x^2 + 2$ and $z = 6 - 2x^2 - y^2$. These surfaces are plotted in Figure 13.45(a) and (b), respectively; the region D is shown in part (c) of the figure.

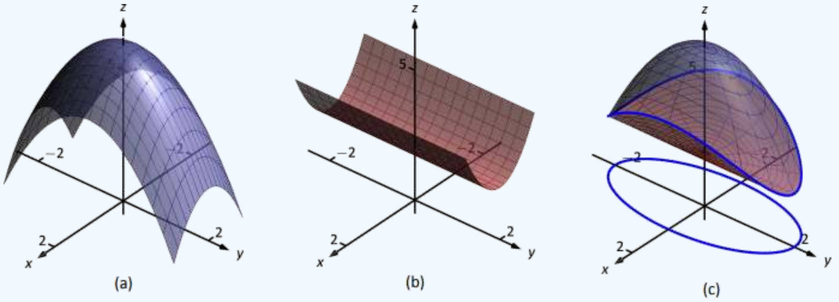


Figure 13.45: The region D is bounded by the surfaces shown in (a) and (b); D is shown in (c).

Solution
The main point of this example is this: integrating with respect to z first is rather straightforward; integrating with respect to x first is not.

The order $dz\,dy\,dx$:

The bounds on z are clearly $2x^2 + 2 \le z \le 6 - 2x^2 - y^2$. Collapsing D onto the x-y plane gives the ellipse shown in Figure 13.45(c). The equation of this ellipse is found by setting the two surfaces equal to each other:

$$2x^2 + 2 = 6 - 2x^2 - y^2 \quad \Rightarrow \quad 4x^2 + y^2 = 4 \quad \Rightarrow \quad x^2 + \frac{y^2}{4} = 1. \tag{13.6.27}$$

We can describe this ellipse with the bounds

$$-\sqrt{4 - 4x^2} \le y \le \sqrt{4 - 4x^2} \quad \text{and} \quad -1 \le x \le 1. \tag{13.6.28}$$

Thus we find volume as

$$\begin{array}{c} 2x^2 + 2 \le z \le 6 - 2x^2 - y^2 \\ -\sqrt{4 - 4x^2} \le y \le \sqrt{4 - 4x^2} \\ -1 \le x \le 1 \end{array} \quad \Rightarrow \int_{-1}^{1}\int_{-\sqrt{4-4x^2}}^{\sqrt{4-4x^2}}\int_{2x^2+2}^{6-2x^2-y^2} dz\,dy\,dx\ . \tag{13.6.29}$$

The order $dydzdx$:

Integrating with respect to y is not too difficult. Since the surface $z = 2x^2 + 2$ is a cylinder whose directrix is the y-axis, it does not create a border for y. The paraboloid $z = 6 - 2x^2 - y^2$ does; solving for y, we get the bounds

$$-\sqrt{6 - 2x^2 - z} \le y \le \sqrt{6 - 2x^2 - z}. \tag{13.6.30}$$

Collapsing D onto the x-z axes gives the region shown in Figure 13.46(a); the lower curve is from the cylinder, with equation $z = 2x^2 + 2$. The upper curve is from the paraboloid; with $y = 0$, the curve is $z = 6 - 2x^2$. Thus bounds on z are $2x^2 + 2 \le z \le 6 - 2x^2$; the bounds on x are $-1 \le x \le 1$. Thus we have:

$$\begin{array}{c} -\sqrt{6-2x^2-z} \le y \le \sqrt{6-2x^2-z} \\ 2x^2+2 \le z \le 6-2x^2 \\ -1 \le x \le 1 \end{array} \qquad \Rightarrow \int_{-1}^{1}\int_{2x^2+2}^{6-2x^2}\int_{-\sqrt{6-2x^2-z}}^{\sqrt{6-2x^2-z}} dy\,dz\,dx. \tag{13.6.31}$$

The order $dx\,dz\,dy$:

This order takes more effort as D must be split into two subregions. The two surfaces create two sets of upper/lower bounds in terms of x; the cylinder creates bounds

$$-\sqrt{z/2-1} \le x \le \sqrt{z/2-1} \tag{13.6.32}$$

for region D_1 and the paraboloid creates bounds

$$-\sqrt{3-y^2/2-z^2/2} \le x \le \sqrt{3-y^2/2-z^2/2} \tag{13.6.33}$$

for region D_2.

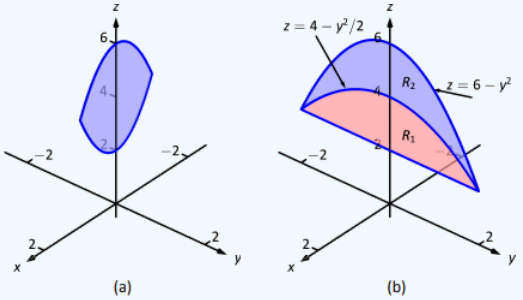


Figure 13.46: The region D in Example 13.6.6 is collapsed onto the x-z plane in (a); in (b), it is collapsed onto the y-z plane.

Collapsing D onto the y-z axes gives the regions shown in Figure 13.46(b). We find the equation of the curve $z=4-y^2/2$ by noting that the equation of the ellipse seen in Figure 13.45(c) has equation

$$x^2+y^2/4=1 \quad\Rightarrow\quad x=\sqrt{1-y^2/4}. \tag{13.6.34}$$

Substitute this expression for x in either surface equation, $z=6-2x^2-y^2$ or $z=2x^2+2$. In both cases, we find

$$z=4-\frac{1}{2}y^2. \tag{13.6.35}$$

Region R_1, corresponding to D_1, has bounds

$$2 \le z \le 4-y^2/2, \quad -2 \le y \le 2 \tag{13.6.36}$$

and region R_2, corresponding to D_2, has bounds

$$4-y^2/2 \le z \le 6-y^2, \quad -2 \le y \le 2. \tag{13.6.37}$$

Thus the volume of D is given by:

$$\int_{-2}^2\int_2^{4-y^2/2}\int_{-\sqrt{z/2-1}}^{\sqrt{z/2-1}} dx \, dz \, dy \ +\ \int_{-2}^2\int_{4-y^2/2}^{6-y^2}\int_{-\sqrt{3-y^2/2-z^2/2}}^{\sqrt{3-y^2/2-z^2/2}} dx \, dz \, dy.$$

If all one wanted to do in Example 13.6.6 was find the volume of the region D, one would have likely stopped at the first integration setup (with order $dz\,dy\,dx$) and computed the volume from there. However, we included the other two methods 1) to show that it could be done, "messy" or not, and 2) because sometimes we "have" to use a less desirable order of integration in order to actually integrate.

Triple Integration and Functions of Three Variables

There are uses for triple integration beyond merely finding volume, just as there are uses for integration beyond "area under the curve." These uses start with understanding how to integrate functions of three variables, which is effectively no different than integrating functions of two variables. This leads us to a definition, followed by an example.

Definition 107 Iterated Integration, (Part II)

Let D be a closed, bounded region in space, over which $g_1(x)$, $g_2(x)$, $f_1(x,y)$, $f_2(x,y)$ and $h(x,y,z)$ are all continuous, and let a and b be real numbers.

The **iterated integral** $\int_a^b\int_{g_1(x)}^{g_2(x)}\int_{f_1(x,y)}^{f_2(x,y)} h(x,y,z)dz\,dy\,dx$ is evaluated as

$$\int_a^b\int_{g_1(x)}^{g_2(x)}\int_{f_1(x,y)}^{f_2(x,y)} h(x,y,z) dz \, dy \, dx = \int_a^b\int_{g_1(x)}^{g_2(x)}\left(\int_{f_1(x,y)}^{f_2(x,y)} h(x,y,z) dz\right) dy \, dx.\]

Example 13.6.7: Evaluating a triple integral of a function of three variables

Evaluate $\int_0^1\int_{x^2}^{x}\int_{x^2-y}^{2x+3y}\left(xy+2xz\right)dz\,dy\,dx$.

Solution
We evaluate this integral according to Definition 107.

$$\begin{aligned}
&\int_0^1\int_{x^2}^{x}\int_{x^2-y}^{2x+3y}\left(xy+2xz\right)dz\,dy\,dx\\
&=\int_0^1\int_{x^2}^{x}\left(\int_{x^2-y}^{2x+3y}\left(xy+2xz\right)dz\right)dy\,dx\\
&=\int_0^1\int_{x^2}^{x}\left(\left(xyz+xz^2\right)\Big|_{x^2-y}^{2x+3y}\right)dy\,dx\\
&=\int_0^1\int_{x^2}^{x}\left(xy(2x+3y)+x(2x+3y)^2-\left(xy(x^2-y)+x(x^2-y)^2\right)\right)dy\,dx\\
&=\int_0^1\int_{x^2}^{x}\left(-x^5+x^3y+4x^3+14x^2y+12xy^2\right)dy\,dx.
\end{aligned}$$

We continue as we have in the past, showing fewer steps.

$$\begin{aligned}
&=\int_0^1\left(-\frac{7}{2}x^7-8x^6-\frac{7}{2}x^5+15x^4\right)dx\\
&=\frac{281}{336}\approx 0.836.
\end{aligned}$$

We now know *how* to evaluate a triple integral of a function of three variables; we do not yet understand what it *means*. We build up this understanding in a way very similar to how we have understood integration and double integration.

Let $h(x,y,z)$ a continuous function of three variables, defined over some space region D. We can partition D into n rectangular--solid subregions, each with dimensions $\Delta x_i\times\Delta y_i\times\Delta z_i$. Let (x_i,y_i,z_i) be some point in the $i^{\text{ th}}$ subregion, and consider the product $h(x_i,y_i,z_i)\Delta x_i\Delta y_i\Delta z_i$. It is the product of a function value (that's the $h(x_i,y_i,z_i)$ part) and a small volume ΔV_i (that's the $\Delta x_i\Delta y_i\Delta z_i$ part). One of the simplest understanding of this type of product is when h describes the density of an object, for then $h\times\text{volume}=\text{mass}$.

We can sum up all n products over D. Again letting $||\Delta D||$ represent the length of the longest diagonal of the n rectangular solids in the partition, we can take the limit of the sums of products as $||\Delta D||\to 0$. That is, we can find
$$ S = \lim_{||\Delta D||\to 0} \sum_{i=1}^n h(x_i,y_i,z_i)\Delta V_i=\lim_{||\Delta D||\to 0} \sum_{i=1}^n h(x_i,y_i,z_i)\Delta x_i\Delta y_i\Delta z_i.\]

While this limit has lots of interpretations depending on the function h, in the case where h describes density, S is the total mass of the object described by the region D.

We now use the above limit to define the **triple integral**, give a theorem that relates triple integrals to iterated iteration, followed by the application of triple integrals to find the centers of mass of solid objects.

Definition 108 Triple Integral

Let $w=h(x,y,z)$ be a continuous function over a closed, bounded space region D, and let ΔD be any partition of D into n rectangular solids with volume ΔV_i. The **triple integral of** h **over** D is
$$\iiint_Dh(x,y,z) dV = \lim_{||\Delta D||\to 0}\sum_{i=1}^n h(x_i,y_i,z_i)\Delta V_i.\]

Note: We previously showed how the summation of rectangles over a region R in the plane could be viewed as a double sum, leading to the double integral. Likewise, we can view the sum $\sum_{i=1}^{n} h(x_i, y_i, z_i)\Delta x_i \Delta y_i \Delta z_i$ as a triple sum,

$$\sum_{k=1}^{p}\sum_{j=1}^{n}\sum_{i=1}^{m} h(x_i, y_j, z_k)\Delta x_i \Delta y_j \Delta z_k, \tag{13.6.38}$$

which we evaluate as

$$\sum_{k=1}^{p}\left(\sum_{j=1}^{n}\left(\sum_{i=1}^{m} h(x_i, y_j, z_k)\Delta x_i\right)\Delta y_j\right)\Delta z_k. \tag{13.6.39}$$

Here we fix a k value, which establishes the z-height of the rectangular solids on one "level" of all the rectangular solids in the space region D. The inner double summation adds up all the volumes of the rectangular solids on this level, while the outer summation adds up the volumes of each level.

This triple summation understanding leads to the $\iiint_D$ notation of the triple integral, as well as the method of evaluation shown in Theorem 127.

The following theorem assures us that the above limit exists for continuous functions h and gives us a method of evaluating the limit.

theorem 127 Triple Integration (Part II)

Let $w = h(x, y, z)$ be a continuous function over a closed, bounded space region D, and let ΔD be any partition of D into n rectangular solids with volume V_i.

1. The limit $\lim_{||\Delta D||\to 0}\sum_{i=1}^{n} h(x_i, y_i, z_i)\Delta V_i$ exists.
2. If D is defined as the region bounded by the planes $x = a$ and $x = b$, the cylinders $y = g_1(x)$ and $y = g_2(x)$, and the surfaces $z = f_1(x, y)$ and $z = f_2(x, y)$, where $a < b$, $g_1(x) \le g_2(x)$ and $f_1(x, y) \le f_2(x, y)$ on D, then

$$\iiint_D h(x, y, z)\,dV = \int_a^b \int_{g_1(x)}^{g_2(x)} \int_{f_1(x,y)}^{f_2(x,y)} h(x, y, z)\,dz\,dy\,dx. \tag{13.6.40}$$

We now apply triple integration to find the centers of mass of solid objects.

Mass and Center of Mass

One may wish to review Section 13.4 for a reminder of the relevant terms and concepts.

Definition 109 Mass, Center of Mass of Solids

Let a solid be represented by a region D in space with variable density function $\delta(x, y, z)$.

1. The **mass** of the object is $M = \iiint_D \,dm = \iiint_D \delta(x, y, z)\,dV$.
2. The **moment about the x-y plane** is $M_{xy} = \iiint_D z\delta(x, y, z)\,dV$.
3. The **moment about the x-z plane** is $M_{xz} = \iiint_D y\delta(x, y, z)\,dV$.
4. The **moment about the y-z plane** is $M_{yz} = \iiint_D x\delta(x, y, z)\,dV$.
5. The **center of mass** of the object is

$$(\overline{x}, \overline{y}, \overline{z}) = \left(\frac{M_{yz}}{M}, \frac{M_{xz}}{M}, \frac{M_{xy}}{M}\right). \tag{13.6.41}$$

Example 13.6.8: Finding the center of mass of a solid

Find the mass and center of mass of the solid represented by the space region bounded by the coordinate planes and $z = 2 - y/3 - 2x/3$, shown in Figure 13.47, with constant density $\delta(x, y, z) = 3\text{gm/cm}^3$. (Note: this space region was used in Example 13.6.2.)

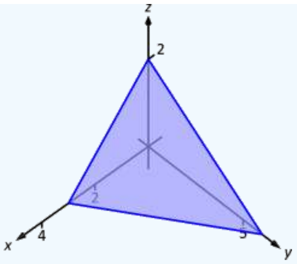


Figure 13.47: Finding the center of mass of this solid in Example 13.6.8.

Solution

We apply Definition 109. In Example 13.6.2, we found bounds for the order of integration $dz\,dy\,dx$ to be $0 \leq z \leq 2 - y/3 - 2x/3$, $0 \leq y \leq 6 - 2x$ and $0 \leq x \leq 3$. We find the mass of the object:

$$\begin{aligned} M &= \iiint_D \delta(x,y,z)dV \\ &= \int_0^3 \int_0^{6-2x} \int_0^{2-y/3-2x/3} \big(3\big)dz\,dy\,dx \\ &= 3\int_0^3 \int_0^{6-2x} \int_0^{2-y/3-2x/3} dz\,dy\,dx \\ &= 3(6) = 18\text{gm}. \end{aligned}$$

The evaluation of the triple integral is done in Example 13.6.2, so we skipped those steps above. Note how the mass of an object with constant density is simply "density×volume."

We now find the moments about the planes.

$$\begin{aligned} M_{xy} &= \iiint_D 3z dV \\ &= \int_0^3 \int_0^{6-2x} \int_0^{2-y/3-2x/3} \big(3z\big)dz\,dy\,dx \\ &= \int_0^3 \int_0^{6-2x} \frac{3}{2}\big(2 - y/3 - 2x/3\big)^2 dy\,dx \\ &= \int_0^3 -\frac{4}{9}\big(x-3\big)^3 dx \\ &= 9. \end{aligned}$$

We omit the steps of integrating to find the other moments.

$$\begin{aligned} M_{yz} &= \iiint_D 3x dV \\ &= \frac{27}{2}. \\ M_{xz} &= \iiint_D 3y dV \\ &= 27. \end{aligned}$$

The center of mass is $$\big(\overline{x},\overline{y},\overline{z}\big) = \left(\frac{27/2}{18},\frac{27}{18},\frac{9}{18}\right) = \big(0.75,1.5,0.5\big).\]

Example 13.6.9: Finding the center of mass of a solid

Find the center of mass of the solid represented by the region bounded by the planes $z = 0$ and $z = -y$ and the cylinder $x^2 + y^2 = 1$, shown in Figure 13.48, with density function $\delta(x,y,z) = 10 + x^2 + 5y - 5z$. (Note: this space region was used in Example 13.6.3.)

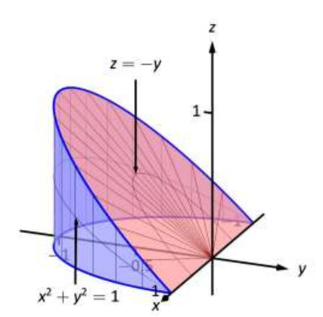


Figure 13.48: Finding the center of mass of this solid in Example 13.6.9.

Solution

As we start, consider the density function. It is symmetric about the y-z plane, and the farther one moves from this plane, the denser the object is. The symmetry indicates that $\overline{x}$ should be 0.

As one moves away from the origin in the y or z directions, the object becomes less dense, though there is more volume in these regions.

Though none of the integrals needed to compute the center of mass are particularly hard, they do require a number of steps. We emphasize here the importance of knowing how to set up the proper integrals; in complex situations we can appeal to technology for a good approximation, if not the exact answer. We use the order of integration $dz\,dy\,dx$, using the bounds found in Example 13.6.4. (As these are the same for all four triple integrals, we explicitly show the bounds only for M.)

$$\begin{aligned}
M &= \iiint_D \big(10+x^2+5y-5z\big)dV\\
&= \int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{0}\int_0^{-y}\big(10+x^2+5y-5z\big)dV\\
&= \frac{64}{5}-\frac{15\pi}{16}\approx 3.855.\\
M_{yz} &= \iiint_D x\big(10+x^2+5y-5z\big)dV\\
&= 0.\\
M_{xz} &= \iiint_D y\big(10+x^2+5y-5z\big)dV\\
&= 2-\frac{61\pi}{48}\approx -1.99.\\
M_{xy} &= \iiint_D z\big(10+x^2+5y-5z\big)dV\\
&= \frac{61\pi}{96}-\frac{10}{9}\approx 0.885.
\end{aligned}$$

Note how $M_{yz}=0$, as expected. The center of mass is
$$\big(\overline{x},\overline{y},\overline{z}\big) = \left(0,\frac{-1.99}{3.855},\frac{0.885}{3.855}\right) \approx \big(0,-0.516, 0.230\big).\]

As stated before, there are many uses for triple integration beyond finding volume. When $h(x,y,z)$ describes a rate of change function over some space region D, then $\iiint_D h(x,y,z)dV$ gives the total change over D. Our one specific example of this was computing mass; a density function is simply a "rate of mass change per volume" function. Integrating density gives total mass.

While knowing how to *integrate* is important, it is arguably much more important to know *how to set up* integrals. It takes skill to create a formula that describes a desired quantity; modern technology is very useful in evaluating these formulas quickly and accurately.

This chapter investigated the natural follow--on to partial derivatives: iterated integration. We learned how to use the bounds of a double integral to describe a region in the plane using both rectangular and polar coordinates, then later expanded to use the bounds of a triple integral to describe a region in space. We used double integrals to find volumes under surfaces, surface area, and the center of mass of lamina; we used triple integrals as an alternate method of finding volumes of space regions and also to find the center of mass of a region in space.

Integration does not stop here. We could continue to iterate our integrals, next investigating "quadruple integrals" whose bounds describe a region in 4--dimensional space (which are very hard to visualize). We can also look back to "regular" integration where we found the area under a curve in the plane. A natural analogue to this is finding the "area under a curve," where the curve is in space, not in a plane. These are just two of many avenues to explore under the heading of "integration."

This page titled 13.6: Volume Between Surfaces and Triple Integration is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

13.E: Applications of Multiple Integration (Exercises)

13.1: Iterated Integrals and Area

Terms and Concepts

1. When integrating $f_x(x,y)$ with respect to x, the constant of integration C is really which: $C(x)$ or $C(y)$? What does this mean?

2. Integrating an integral is called _________ __________.

3. When evaluating an iterated integral, we integrate from _______ to ________, then from _________ to __________.

4. One understanding of an iterated integral is that $\int_a^b \int_{g_1(x)}^{g_2(x)} dy\,dx$ gives the _______ of a plane region.

Problems

In Exercises 5-10, evaluate the integral and subsequent iterated integral.

5.

(a) $\int_2^5 (6x^2+4xy-3y^2)\,dy$

(b) $\int_{-3}^2 \int_2^5 (6x^2+4xy-3y^2)\,dy\,dx$

6.

(a) $\int_0^\pi (2x\cos y+\sin x)\,dx$

(b) $\int_0^{\pi/2} \int_0^\pi (2x\cos y+\sin x)\,dx\,dy$

7.

(a) $\int_1^x (x^2y-y+2)\,dy$

(b) $\int_0^2 \int_1^x (x^2y-y+2)\,dy\,dx$

8.

(a) $\int_y^{y^2} (x-y)\,dx$

(b) $\int_{-1}^1 \int_y^{y^2} (x-y)\,dx\,dy$

9.

(a) $\int_0^y (\cos x\sin y)\,dx$

(b) $\int_0^\pi \int_0^y (\cos x\sin y)\,dx\,dy$

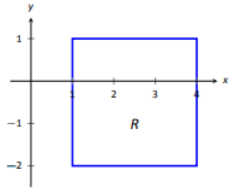
10.

(a) $\int_0^x \left(\frac{1}{1+x^2}\right)\,dy$

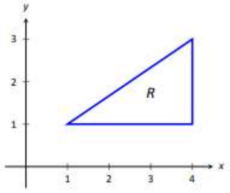
(b) $\int_1^2 \int_0^x \left(\frac{1}{1+x^2}\right)\,dy\,dx$

In Exercises 11-16, a graph of a planar region R is given. Give the iterated integrals, with both orders of integration $dy\,dx$ and $dx\,dy$, that give the area of R. Evaluate one of the iterated integrals to find the area.

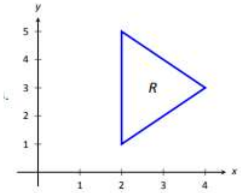
11.



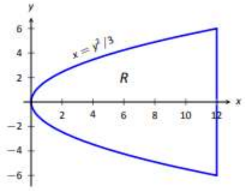
12.


y
3
2
1
R
1
2
3
4
x

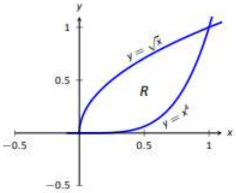
13.


y
5
4
3
2
1
R
1
2
3
4
x

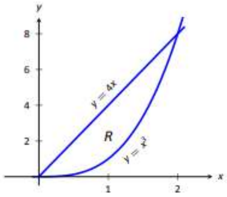
14.


y
6
4
2
−2
−4
−6
x = y^2/3
R
2
4
6
8
10
12
x

15.


y
1
0.5
−0.5
y = √x
R
y = x^4
−0.5
0.5
1
x

16.


y
8
6
4
2
y = 4x
R
y = x^3
1
2
x

In Exercises 17-22, iterated integrals are given that compute the area of a region *R* in the xy-plane. Sketch the region *R*, and give the iterated integral(s) that give the area of *R* with the opposite order of integration.

17. $\int_{-2}^{2}\int_{0}^{4-x^2} dy\,dx$

18. $\int_{0}^{1}\int_{5-5x}^{5-5x^2} dy\,dx$

19. $\int_{-2}^{2}\int_{0}^{2\sqrt{4-y^2}} dx\,dy$

20. $\int_{-3}^{3}\int_{-\sqrt{9-x^2}}^{\sqrt{9-x^2}} dy\,dx$

21. $\int_0^1\int_{-\sqrt{y}}^{\sqrt{y}} dx\,dy+\int_1^4\int_{y-2}^{\sqrt{y}} dx\,dy$

22. $\int_{-1}^1\int_{(x-1)/2}^{(1-x)/2} dy\,dx$

13.2: Double Integration and Volume

Terms and Concepts

1. An integral can be interpreted as giving the signed area over an interval; a double integral can be interpreted as giving the signed ________ over a region.

2. Explain why the following statement is false: "Fubini's Theorem states that $\int_a^b\int_{g_1(x)}^{g_2(x)} f(x,y)\,dy\,dx=\int_a^b\int_{g_1(y)}^{g_2(y)} f(x,y)\,dx\,dy$."

3. Explain why if $f(x,y)>0$ over a region R, then $\int\int_R f(x,y)\,dA>0$.

4. If $\int\int_R f(x,y)dA=\int\int_R g(x,y)\,dA$, does this imply $f(x,y)=g(x,y)$?

Problems

In Exercises 5-10,
(a) Evaluate the given iterated integral, and
(b) rewrite the integral using the other order of integration.

5. $\int_1^2\int_{-1}^1\left(\frac{x}{y}+3\right)\,dx\,dy$

6. $\int_{-\pi/2}^{\pi/2}\int_0^{\pi}(\sin x\cos y$\,dy\,dx\)

7. $\int_0^4\int_0^{-x/2+2}\left(3x^2-y+2\right)\,dy\,dx$

8. $\int_1^3\int_y^3\left(x^2y-xy^2\right)\,dx\,dy$

9. $\int_0^2 1\int_{-\sqrt{1-y}}^{\sqrt{1-y}}(x+y+2)\,dx\,dy$

10. $\int_0^9\int_{y/3}^{\sqrt{3}}\left(xy^2\right)\,dx\,dy$

In Exercises 11-18:
(a) Sketch the region *R* given by the problem.
(b) Set up the iterated integrals, in both orders, that evaluate the given double integral for the described region *R*.
(c) Evaluate one of the iterated integrals to find the signed volume under the surface $z=f(x,y)$ over the region *R*.

11. $\int\int_R x^2y\,dA$, where R is bounded by $y=\sqrt{x}$ and $y=x^2$.

12. $\int\int_R x^2y\,dA$, where R is bounded by $y=\sqrt[3]{x}$ and $y=x^3$.

13. $\int\int_R x^2-y^2\,dA$, where R is the rectangle with corners $(-1,-1),(1,-1),(1,1)$ and $(-1,1)$.

14. $\int\int_R ye^x\,dA$, where R is bounded by $x=0,\ x=y^2$ and $y=1$.

15. $\int\int_R(6-3x-2y)\,dA$, where R is bounded by $x=0,y=0$ and $3x+2y=6$.

16. $\int\int_R e^y\,dA$, where R is bounded by $y=\ln x$ and $y=\frac{1}{e-1}(x-1)$.

17. $\int\int_R(x^3y-x)\,dA$, where R is the half of the circle $x^2+y^2=9$ in the first and second quadrants.

18. $\int\int_R(4-sy)\,dA$, where R is bounded by $y=0,y=x/e$ and $y=\ln x$.

In Exercises 19-22, state why it is difficult/impossible to integrate the iterated integral in the given order of integration. Change the order of integration and evaluate the new iterated integral.

19. $\int_0^4\int_{y/2}^2 e^{x^2}\,dx\,dy$

20. $\int_0^{\sqrt{\pi/2}}\int_x^{\sqrt{\pi/2}}\cos(y^2)\,dy\,dx$

21. $\int_0^1\int_y^1\frac{2y}{x^2+y^2}\,dx\,dy$

22. $\int_{-1}^1\int_1^2\frac{x\tan^2 y}{1+\ln y}\,dy\,dx$

In Exercises 23-26, find the average value of *f* over the region *R*. Notice how these functions and regions are related to the iterated integrals given in Exercises 5-8.

23. $f(x,y) = \frac{x}{y} + 3$; *R* is the rectangle with opposite corners $(-1,1)$ and $(1,2)$.

24. $f(x,y) = \sin x \cos y$; *R* is bounded by $x = 0, x = \pi, y = -\pi/2$ and $y = \pi/2$.

25. $f(x,y) = 3x^2 - y + 2$; *R* is bounded by the lines $y = 0, y = 2 - x/2$ and $x = 0$.

26. $f(x,y) = x^2y - xy^2$; *R* is bounded by $y = x, y = 1$ and $x = 3$.

13.3: Double Integration with Polar Coordinates

Terms and Concepts

1. When evaluating $\int\int_R f(x,y)\,dA$ using polar coordinates, $f(x,y)$ is replaced with _______ and dA is replaced with _______.

2. Why would one be interested in evaluating a double integral with polar coordinates?

Problems

In Exercises 3-10, a function $f(x,y)$ is given and a region *R* of the *x*-*y* plane is described. Set up and evaluate $\int\int_R f(x,y)\,dA$.

3. $f(x,y) = 3x - y + 4$; *R* is the region enclosed by the circle $x^2 + y^2 = 1$.

4. $f(x,y) = 4x + 4y$; *R* is the region enclosed by the circle $x^2 + y^2 = 4$.

5. $f(x,y) = 8 - y$; *R* is the region enclosed by the circles with polar equations $r = \cos\theta$ and $r = 3\cos\theta$.

6. $f(x,y) = 4$; *R* is the region enclosed by the petal of the rose curve $r = \sin(2\theta)$ in the first quadrant.

7. $f(x,y) = \ln(x^2 + y^2)$; *R* is the annulus enclosed by the circles \(x^2+y^2=1\text{ and }x^2+y^2=4.

8. $f(x,y) = 1 - x^2 - y^2$; *R* is the region enclosed by the circle $x^2 + y^2 = 1$.

9. $f(x,y) = x^2 - y^2$; *R* is the region enclosed by the circle $x^2 + y^2 = 36$ in the first and fourth quadrants.

10. $f(x,y) = (x - y)/(x + y)$; *R* is the region enclosed by the lines $y = x, y = 0$ and the circle $x^2 + y^2 = 1$ in the first quadrant.

In Exercises 11-14, an iterated integral in rectangular coordinates is given. Rewrite the integral using polar coordinates and evaluate the new double integral.

11. $\int_0^5 \int_{-\sqrt{25-x^2}}^{\sqrt{25-x^2}} \sqrt{x^2+y^2}\,dy\,dx$

12. $\int_{-4}^4 \int_{-\sqrt{16-y^2}}^0 (2y - x)dx\ dy$

13. $\int_0^2 \int_y^{\sqrt{8-y^2}} (x + y)\,dx\,dy$

14. $\int_{-2}^{-1} \int_0^{\sqrt{4-x^2}} (x+5)dy\,dx + \int_{-1}^1 \int_{\sqrt{1-x^2}}^{\sqrt{4-x^2}} (x+5)\,dy\,dx + \int_1^2 \int_0^{\sqrt{4-x^2}} (x+5)\,dy\,dx$

In Exercises 15-16, special double integrals are presented that are especially well suited for evaluation in polar coordinates.

15. Consider $\int\int_R e^{-(x^2+y^2)}dA$.
(a) Why is this integral difficult to evaluate in rectangular coordinates, regardless of the region *R*?
(b) Let *R* be the region bounded by the circle of radius *a* centered at the origin. Evaluate the double integral using polar coordinates.
(c) Take the limit of your answer from (b), as $a \to \infty$. What does this imply about the volume under the surface of $e^{-(x^2+y^2)}$ over the entire *x*-*y* plane?

16. The surface of a right circular cone with height *h* and base radius *a* can be described by the equation $f(x,y) = h - h\sqrt{\frac{x^2}{a^2} + \frac{y^2}{a^2}}$, where the tip of the cone lies at $(0,0,h)$ and the circular base lies in the *x*-*y* plane, centered at the origin.
Confirm that the volume of a right circular cone with height *h* and base radius *a* is $V = \frac{1}{3}\pi a^2 h$ by evaluating $\int\int_R f(x,y)\,dA$ in polar coordinates.

13.4: Center of Mass

Terms and Concepts

1. Why is it easy to use "mass" and "weight" interchangeably, even though they are different measures?

2. Given a point (x, y), the value of *x* is a measure of distance from the _________-axis.

3. We can think of $\int\int_R dm$ as meaning "sum up lots of ________."

4. What is a "discrete planar system?"

5. Why does M_x use $\int\int_R y\delta(x,y)\,dA$ instead of $\int\int_R x\delta(x,y)\,dA$; that is, why do we use "y" and not "x"?

6. Describe a situation where the center of mass of a lamina does not lie within the region of the lamina itself.

Problems

In Exercises 7-10, point masses are given along a line or in the plane. Find the center of mass $\overline{x}$ or $(\overline{x}, \overline{y})$, as appropriate. (All masses are in grams and distances are in cm.)

7. $m_1 = 4 \text{ at } x = 1; \quad m_2 = 3 \text{ at } x = 3; \quad m_3 = 5 \text{ at } x = 10$

8. $m_1 = 2 \text{ at } x = -3; \quad m_2 = 2 \text{ at } x = -1; \quad m_3 = 3 \text{ at } x = 0; \quad m_4 = 3 \text{ at } x = 7$

9. $m_1 = 2 \text{ at } (-2, 2); \quad m_2 = 2 \text{ at } (2, -2); \quad m_3 = 20 \text{ at } (0, 4)$

10. $m_1 = 1 \text{ at } (-1, 1); \quad m_2 = 2 \text{ at } (-1, 1); \quad m_3 = 2 \text{ at } (1, 1); \quad m_4 = 1 \text{ at } (1, -1)$

In Exercises 11-18, find the mass/weight of the lamina described by the region *R* in the plane and its density function $\delta(x, y)$.

11. *R* is the rectangle with corners $(1, -3), (1, 2), (7, 2) \text{ and } (7, -3); \delta(x, y) = 5$gm/cm^2

12. *R* is the rectangle with corners $(1, -3), (1, 2), (7, 2) \text{ and } (7, -3); \delta(x, y) = (x + y^2)$gm/cm^2

13. *R* is the triangle with corners $(-1, 0), (1, 0), \text{ and } (0, 1); \delta(x, y) = 2$lb/in^2

14. *R* is the triangle with corners $(0, 0), (1, 0), \text{ and } (0, 1); \delta(x, y) = (x^2 + y^2 + 1)$lb/in^2

15. *R* is the circle centered at the origin with radius 2; $\delta(x, y) = (x + y + 4)$kg/m^2

16. *R* is the circle sector bounded by $x^2 + y^2 = 25$ in the first quadrant; $\delta(x, y) = (\sqrt{x^2 + y^2} + 1)$kg/m^2

17. *R* is the annulus in the first and second quadrants bounded by $x^2 + y^2 = 9 \text{ and } x^2 + y^2 = 36; \delta(x, y) = 4$lb/ft^2

18. *R* is the annulus in the first and second quadrants bounded by $x^2 + y^2 = 9 \text{ and } x^+ y^2 = 36; \delta(x, y) = \sqrt{x^2 + y^2}$lb/ft^2

In Exercises 19-26, find the center of mass of the lamina described by the region *R* in the plane and its density function $\delta(x, y)$.

Note: these are the same lamina as in Exercises 11-18.

19. *R* is the rectangle with corners $(1, -3), (1, 2), (7, 2) \text{ and } (7, -3); \delta(x, y) = 5$gm/cm^2

20. *R* is the rectangle with corners $(1, -3), (1, 2), (7, 2) \text{ and } (7, -3); \delta(x, y) = (x + y^2)$gm/cm^2

21. *R* is the triangle with corners $(-1, 0), (1, 0), \text{ and } (0, 1); \delta(x, y) = 2$lb/in^2

22. *R* is the triangle with corners $(0, 0), (1, 0), \text{ and } (0, 1); \delta(x, y) = (x^2 + y^2 + 1)$lb/in^2

23. *R* is the circle centered at the origin with radius 2; $\delta(x, y) = (x + y + 4)$kg/m^2

24. *R* is the circle sector bounded by $x^2 + y^2 = 25$ in the first quadrant; $\delta(x, y) = (\sqrt{x^2 + y^2} + 1)$kg/m^2

25. *R* is the annulus in the first and second quadrants bounded by $x^2 + y^2 = 9 \text{ and } x^2 + y^2 = 36; \delta(x, y) = 4$lb/ft^2

26. *R* is the annulus in the first and second quadrants bounded by $x^2 + y^2 = 9 \text{ and } x^+ y^2 = 36; \delta(x, y) = \sqrt{x^2 + y^2}$lb/ft^2

The moment of inertia i is a measure of the tendency of lamina to resist rotating about an axis or continue to rotate about an axis. i_x is the moment of inertia about the x-axis, i_x is the moment of inertia about the x-axis, and i_o is the moment of inertia about the origin.These are computed as follows:

- $i_x = \int\int_R y^2\,dm$
- $i_y = \int\int_R x^2\,dm$
- $i_o = \int\int_R (x^2 + y^2)\,dm$

In Exercises 27-30, a lamina corresponding to a planar region *R* is given with a mass of 16 units. For each, compute i_x, i_y and i_o.

27. *R* is the 4 x 4 square with corners $(-2,-2)$ and $(2,2)$ with density $\delta(x,y)=1$.

28. *R* is the 8 x 2 rectangle with corners $(-4,-1)$ and $(4,1)$ with density $\delta(x,y)=1$.

29. *R* is the 4 x 2 rectangle with corners $(-2,-1)$ and $(2,1)$ with density $\delta(x,y)=2$.

30. *R* is the circle with radius 2 centered at the origin with density $\delta(x,y)=4/\pi$.

13.5: Surface Area

Terms and Concepts

1. "Surface area" is analogous to what previously studied concept?

2. To approximate the area of a small portion of a surface, we computed the area of its ______ plane.

3. We interpret $\int\int_R \, dS$ as "sum up lots of little _______ ________."

4. Why is it important to know how to set up a double integral to compute surface area, even if the resulting integral is hard to evaluate?

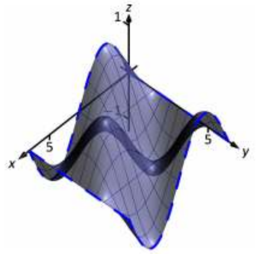
5. Why do $z=f(x,y)$ and $z=g(x,y)=f(x,y)+h$, for some real number *h*, have the same surface area over a region *R*?

6. Let $z=f(x,y)$ and $z=g(x,y)=2f(x,y)$. Why is the surface area of *g* over a region *R* not twice the surface area of f over R?

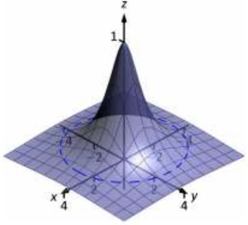
Problems

In Exercises 7-10, set up the iterated integral that computes the surfaces area of the given surface over the region *R*.

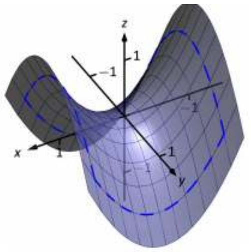
7. $f(x,y)=\sin x\cos y$; R is the rectangle with bounds $0\le x\le 2\pi$, $0\le y\le 2\pi$.



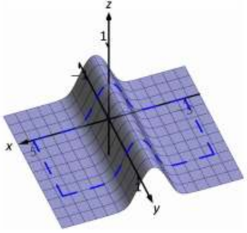
8. $f(x,y)=\frac{1}{x^2+y^2+1}$; R is the circle $x^2+y^2=9$.



9. $f(x,y)=x^2-y^2$; R is the rectangle with opposite corners $(-1,-1)$ and $1,1)$.



10. $f(x,y)=\frac{1}{e^{x^2}+1}$; R is the rectangle bounded by $-5\leq x\leq 5$ and $0\leq y\leq 1$.



In Exercises 11-19, find the area of the given surface over the region *R*.

11. $f(x,y)=3x-7y+2$; R is the rectangle with opposite corners $(-1,0)$ and $(1,3)$.

12. $f(x,y)=2x+2y+2$; R is the triangle with corners $(0,0),(1,0)$ and $(0,1)$.

13. $f(x,y)=x^2+y^2+10$; R is the circle $x^2+y^2=16$.

14. $f(x,y)=-2x+4y^2+7$ over R, the triangle bounded by $y=-x, y=x, 0\leq y\leq 1$.

15. $f(x,y)=x^2+y$ over *R*, the triangle bounded by $y=2x, y=0$ and $x=2$.

16. $f(x,y)=\frac{2}{3}x^{3/2}$ over *R*, the rectangle with opposite corners $(0,0)$ and $(1,1)$.

17. $f(x,y)=10-2\sqrt{x^2+y^2}$ over *R*, the circle $x^2+y^2=25$. (This is the cone with height 10 and base radius 5; be sure to compare your result with the known formula.)

18. Find the surface area of the sphere with radius 5 by doubling the surface area of $f(x,y)=\sqrt{25-x^2-y^2}$ over *R*, the circle $x^2+y^2=25$. (Be sure to compare your result with the known formula.)

19. Find the surface area of the ellipse formed by restricting the plane $f(x,y)=cx+dy+h$ to the region *R*, the circle $x^2+y^2=1$, where *c*, *d* and *h* are some constants. Your answer should be given in terms of *c* and *d*; why does the value of *h* not matter?

13.6: Volume Between Surfaces and Triple Integration

Terms and Concepts

1. The strategy for establishing bounds for triple integrals is "________ to ________, _________ and __________ to _______."

2. Give an informal interpretation of what $"\int\int\int_D dV"$ means.

3. Give two uses of triple integration.

4. If an object has a constant density δ and a volume *V*, what is its mass?

Problems

In Exercises 5-8, two surfaces $f_1(x,y)$ and $f_2(x,y)$ and a region R in the xy-plane are given. Set up and evaluate the triple integral that represents the volume between these surfaces over R.

5. $f_1(x,y)=8-x^2-y^2,\ f_2(x,y)=2x+y$;
R is the square with corners $(-1,-1)$ and $(1,1)$.

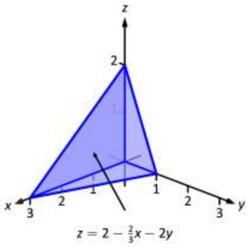
6. $f_1(x,y)=x^2+y^2,\ f_2(x,y)=-x^2-y^2$;
R is the square with corners $(0,0)$ and $(2,3)$.

7. $f_1(x,y)=\sin x\cos y,\ f_2(x,y)=\cos x\sin y+2$;
R is the triangle with corners $(0,0),(\pi,0)$ and (π,π).

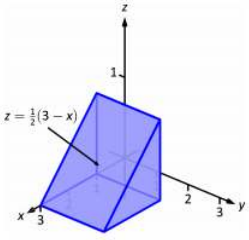
8. $f_1(x,y)=2x^2+2y^2+3,\ f_2(x,y)=6-x^2-y^2$;
R is the circle $x^2+y^2=1$.

In Exercises 9-16, a domain *D* is described by its bounding surfaces, along with a graph. Set up the triple integrals that give the volume of *D* in all 6 orders of integration, and find the volume of *D* by evaluating the indicated triple integral.

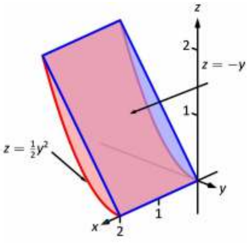
9. *D* is bounded by the coordinate planes and $z=2-2x/3-2y$.
Evaluate the triple integral with order *dz dy dx*.


z
2
1
x
3
2
1
1
2
3
y
$z = 2 - \frac{2}{3}x - 2y$

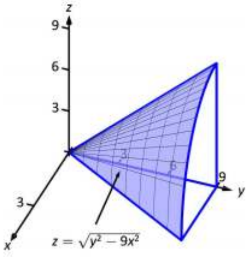
10. *D* is bounded by the planes $y = 0, y = 2, x = 1, z = 0$ and $z = (2 - x)/2$. Evaluate the triple integral with order *dx dy dz*.


z
1
$z = \frac{1}{2}(3 - x)$
x
3
2
3
y

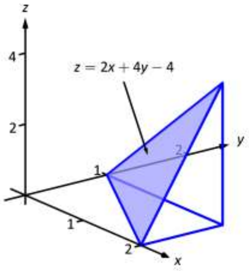
11. *D* is bounded by the planes $x = 0, x = 2, z = -y$ and by $z = y^2/2$. Evaluate the triple integral with order *dy dz dx*.


z
2
$z = -y$
1
$z = \frac{1}{2}y^2$
y
1
x
2

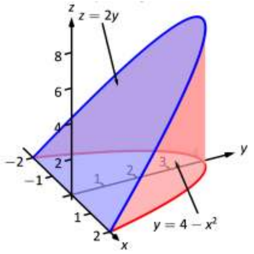
12. *D* is bounded by the planes $z = 0, y = 9, x = 0$ and by $z = \sqrt{y^2 - 9x^2}$. Do not evaluate any triple integral.


z
9
6
3
3
6
9
y
3
x
$z = \sqrt{y^2 - 9x^2}$

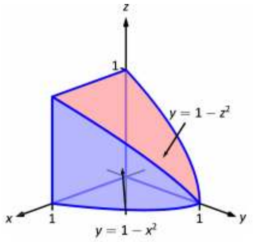
13. *D* is bounded by the planes $x = 2, y = 1, z = 0$ and $z = 2x + 4y - 4$. Evaluate the triple integral with order *dx dy dz*.


z
4
$z = 2x + 4y - 4$
2
y
2
1
1
2
x

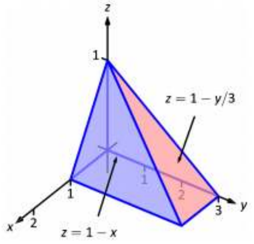
14. *D* is bounded by the plane $z = 2y$ and by $y = 4 - x^2$. Evaluate the triple integral with order *dz dy dx*.



15. *D* is bounded by the coordinate planes and $y = 1 - x^2$ and $y = 1 - z^2$.
Do not evaluate any triple integral. Which order is easier to evaluate: *dz dy dx* or *dy dz dx*? Explain why.



16. *D* is bounded by the coordinate planes and by $z = 1 - y/3$ and $z = 1 - x$.
Evaluate the triple integral with order *dx dy dz*.



In Exercises 17-20, evaluate the triple integral.

17. $\int_{-\pi/2}^{\pi/2} \int_0^{\pi} \int_0^{\pi} (\cos x \sin y \sin z) dz\, dy\, dx$

18. $\int_0^1 \int_0^x \int_0^{x+y} (x + y + z) dz\, dy\, dx$

19. $\int_0^{\pi} \int_0^1 \int_0^z (\sin(yz)) dx\ dy\, dz$

20. $\int_{\pi}^{\pi^2} \int_x^{x^3} \int_{-y^2}^{y^2} (\cos x \sin y \sin z) dz\, dy\, dx$

In Exercises 21-24, find the center of mass of the solid represented by the indicated space region *D* with density function $\delta(x, y, z)$.

21. *D* is bounded by the coordinate planes and $z = 2 - 2x/3 - 2y$; $\delta(x, y, z) = 10\text{g/cm}^3$.
(Note: this is the same region as used in Exercise 9.)

22. *D* is bounded by the planes $y = 0, y = 2, x = 1, z = 0$ and $z = (3 - x)/2$; $\delta(x, y, z) = 2\text{g/cm}^3$.
(Note: this is the same region as used in Exercise 10.)

23. *D* is bounded by the planes $x = 2, y = 1, z = 0$ and $z = 2x + 4y - 4$; $\delta(x, y, z) = x^2\text{lb/in}^3$.
(Note: this is the same region as used in Exercise 13.)

24. *D* is bounded by the planes $z = 2y$ and by $y = 4 - x^2$. $\delta(x, y, z) = y^2\text{lb/in}^3$.
(Note: this is the same region as used in Exercise 14.)

13.E: Applications of Multiple Integration (Exercises) is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

14: Appendix

Topic hierarchy

14.1: Section 1-
14.2: Section 2-
14.3: Section 3-
14.4: Section 4-
14.5: Section 5-
14.6: Section 6-

Contributors and Attributions

- Gregory Hartman (Virginia Military Institute). Contributions were made by Troy Siemers and Dimplekumar Chalishajar of VMI and Brian Heinold of Mount Saint Mary's University. This content is copyrighted by a Creative Commons Attribution - Noncommercial (BY-NC) License. http://www.apexcalculus.com/

This page titled 14: Appendix is shared under a CC BY-NC 3.0 license and was authored, remixed, and/or curated by Gregory Hartman et al. via source content that was edited to the style and standards of the LibreTexts platform.

Welcome to the Mathematics Library. This Living Library is a principal hub of the LibreTexts project, which is a multi-institutional collaborative venture to develop the next generation of open-access texts to improve postsecondary education at all levels of higher learning. The LibreTexts approach is highly collaborative where an Open Access textbook environment is under constant revision by students, faculty, and outside experts to supplant conventional paper-based books.

Welcome to the Mathematics Library. This Living Library is a principal hub of the LibreTexts project, which is a multi-institutional collaborative venture to develop the next generation of open-access texts to improve postsecondary education at all levels of higher learning. The LibreTexts approach is highly collaborative where an Open Access textbook environment is under constant revision by students, faculty, and outside experts to supplant conventional paper-based books.

Welcome to the Mathematics Library. This Living Library is a principal hub of the LibreTexts project, which is a multi-institutional collaborative venture to develop the next generation of open-access texts to improve postsecondary education at all levels of higher learning. The LibreTexts approach is highly collaborative where an Open Access textbook environment is under constant revision by students, faculty, and outside experts to supplant conventional paper-based books.

Welcome to the Mathematics Library. This Living Library is a principal hub of the LibreTexts project, which is a multi-institutional collaborative venture to develop the next generation of open-access texts to improve postsecondary education at all levels of higher learning. The LibreTexts approach is highly collaborative where an Open Access textbook environment is under constant revision by students, faculty, and outside experts to supplant conventional paper-based books.

Welcome to the Mathematics Library. This Living Library is a principal hub of the LibreTexts project, which is a multi-institutional collaborative venture to develop the next generation of open-access texts to improve postsecondary education at all levels of higher learning. The LibreTexts approach is highly collaborative where an Open Access textbook environment is under constant revision by students, faculty, and outside experts to supplant conventional paper-based books.

Welcome to the Mathematics Library. This Living Library is a principal hub of the LibreTexts project, which is a multi-institutional collaborative venture to develop the next generation of open-access texts to improve postsecondary education at all levels of higher learning. The LibreTexts approach is highly collaborative where an Open Access textbook environment is under constant revision by students, faculty, and outside experts to supplant conventional paper-based books.

Index

A
absolute convergence
8.5: Alternating Series and Absolute Convergence
acceleration
11.3: The Calculus of Motion
Alternating Harmonic Series
8.5: Alternating Series and Absolute Convergence
alternating series
8.5: Alternating Series and Absolute Convergence
antiderivative
5.1: Antiderivatives and Indefinite Integration
arc length
7.4: Arc Length and Surface Area
9.5: Calculus and Polar Functions
11.5: The Arc Length Parameter and Curvature
Arc length of a parametric curve
9.3: Calculus and Parametric Equations
arc length parameter
11.5: The Arc Length Parameter and Curvature

B
bisection method
1.5: Continuity
bounds of integration.
5.2: The Definite Integral

C
Cartesian coordinate plane
10.1: Introduction to Cartesian Coordinates in Space
Cartesian Coordinates in Space
10.1: Introduction to Cartesian Coordinates in Space
center of mass
13.4: Center of Mass
chain rule
2.5: The Chain Rule
comparison test
8.3: Integral and Comparison Tests
concavity
3.4: Concavity and the Second Derivative
9.3: Calculus and Parametric Equations
Concavity (parametric equations)
9.3: Calculus and Parametric Equations
conic sections
10.1: Introduction to Cartesian Coordinates in Space
conics
9.1: Conic Sections
constant multiple rule
5.1: Antiderivatives and Indefinite Integration
Constant Rule
2.3: Basic Differentiation Rules
continuity
1.5: Continuity
12.2: Limits and Continuity of Multivariable Functions
contour map
12.1: Introduction to Multivariable Functions
convergence of a series
8.2: Infinite Series
cross product
10.4: The Cross Product
curvature
11.5: The Arc Length Parameter and Curvature
curve sketching
3.5: Curve Sketching
cusps
9.2: Parametric Equations
cylinders
10.1: Introduction to Cartesian Coordinates in Space

D
decreasing on the interval I
3.3: Increasing and Decreasing Functions
definite integral
5.2: The Definite Integral
Derivative of Inverse Functions
2.7: Derivatives of Inverse Functions
Derivative of Inverse Trigonometric Functions
2.7: Derivatives of Inverse Functions
Derivative of logarithmic functions
2.7: Derivatives of Inverse Functions
Difference Rule
5.1: Antiderivatives and Indefinite Integration
Differentiability (two variables)
12.4: Differentiability and the Total Differential
differentials
4.4: Differentials
differentiation rules
2.3: Basic Differentiation Rules
Directional Derivatives
12.6: Directional Derivatives
directrix
9.1: Conic Sections
disk method
7.2: Volume by Cross-Sectional Area- Disk and Washer Methods
distance
10.1: Introduction to Cartesian Coordinates in Space
Distance between a Plane and a Point
10.6: Planes
double integral
13.2: Double Integration and Volume
DOUBLE INTEGRAL IN POLAR COORDINATES
13.3: Double Integration with Polar Coordinates

E
eccentricity
9.1: Conic Sections
elementary function
5.5: Numerical Integration
ellipses
9.1: Conic Sections
extrema
3.1: Extreme Values
12.8: Extreme Values
extreme value theorem
3.1: Extreme Values

F
factorial
8.1: Sequences
FINDING POSITION USING VELOCITY
5.2: The Definite Integral
First Derivative Test
3.3: Increasing and Decreasing Functions
fluids
7.6: Fluid Forces
fundamental theorem of calculus
5.4: The Fundamental Theorem of Calculus

G
generalized chain rule
2.5: The Chain Rule
generalized power rule
2.5: The Chain Rule

H
Higher Order Derivatives
2.3: Basic Differentiation Rules
hyperbolas
9.1: Conic Sections
hyperbolic cosine
6.6: Hyperbolic Functions
hyperbolic functions
6.6: Hyperbolic Functions
hyperbolic sine
6.6: Hyperbolic Functions

I
implicit differentiation
2.6: Implicit Differentiation
4.2: Related Rates
12.5: The Multivariable Chain Rule
improper integral
6.8: Improper Integration
increasing on the interval I
3.3: Increasing and Decreasing Functions
independent variable
5.1: Antiderivatives and Indefinite Integration
indeterminate forms
6.7: L'Hopital's Rule
infinite series
8.2: Infinite Series
inflection point
3.4: Concavity and the Second Derivative
instantaneous velocity
2.1: Instantaneous Rates of Change- The Derivative
3.2: The Mean Value Theorem
Integral Test
8.3: Integral and Comparison Tests
integration by parts
6.2: Integration by Parts
intermediate value theorem
1.5: Continuity
iterated integration
13.1: Iterated Integrals and Area

L
L'Hopital's Rule
6.7: L'Hopital's Rule

lamina
  13.4: Center of Mass
level curve of a function of two variables
  12.1: Introduction to Multivariable Functions
level curves
  12.1: Introduction to Multivariable Functions
level surface of a function of three variables
  12.1: Introduction to Multivariable Functions
limacon
  9.5: Calculus and Polar Functions
limit
  1.2: Epsilon-Delta Definition of a Limit
limits
  1.1: An Introduction to Limits
  1.3: Finding Limits Analytically
Limits of Piecewise functions
  1.4: One Sided Limits
local extremum
  3.1: Extreme Values
logarithmic differentiation
  2.6: Implicit Differentiation

M
maclaurin series
  8.8: Taylor Series
maxima
  4.3: Optimization
mean value theorem
  3.2: The Mean Value Theorem
midpoint rule
  5.3: Riemann Sums
minima
  4.3: Optimization
  12.8: Extreme Values
mixed partial derivatives
  12.3: Partial Derivatives
Multiple Integration
  13.1: Iterated Integrals and Area
Multivariable Chain Rule
  12.5: The Multivariable Chain Rule
multivariable functions
  12.1: Introduction to Multivariable Functions

N
Newton's method
  4.1: Newton's Method
normal lines (parametric equations)
  9.3: Calculus and Parametric Equations
normal vector
  10.6: Planes
  11.4: Unit Tangent and Normal Vectors

O
one sided limit
  1.4: One Sided Limits
optimization
  4.3: Optimization
orthogonal decomposition
  10.3: The Dot Product
orthogonal projection
  10.3: The Dot Product
osculating circle
  11.5: The Arc Length Parameter and Curvature

P
parabola
  9.1: Conic Sections
Parallel Line
  10.5: Lines
Parallelogram law
  10.2: An Introduction to Vectors
parametric equations
  9.2: Parametric Equations
parametric equations of a line
  10.5: Lines
partial derivative
  12.3: Partial Derivatives
partial fraction decomposition
  6.5: Partial Fraction Decomposition
partial sum
  8.2: Infinite Series
perimeter
  4.3: Optimization
plane
  10.1: Introduction to Cartesian Coordinates in Space
  10.6: Planes
polar coordinates
  9.4: Introduction to Polar Coordinates
power rule
  2.3: Basic Differentiation Rules
  2.6: Implicit Differentiation
  5.1: Antiderivatives and Indefinite Integration
power series
  8.6: Power Series
pressure
  7.6: Fluid Forces
projectile motion
  11.3: The Calculus of Motion
projection
  10.3: The Dot Product

Q
quotient rule
  2.4: The Product and Quotient Rules

R
ratio test
  8.4: Ratio and Root Tests
related rates
  4.2: Related Rates
Riemann sums
  5.3: Riemann Sums
Rolle’s Theorem
  3.2: The Mean Value Theorem
root test
  8.4: Ratio and Root Tests

S
saddle point
  12.8: Extreme Values
Second Derivative
  3.4: Concavity and the Second Derivative
second derivative test
  3.4: Concavity and the Second Derivative
  12.8: Extreme Values
second partial derivatives
  12.3: Partial Derivatives
sequences
  8: Sequences and Series
  8.1: Sequences
series
  8: Sequences and Series
shell method
  7.3: The Shell Method
Simpson's rule
  5.5: Numerical Integration
skew lines
  10.5: Lines
solid of revolution
  7.4: Arc Length and Surface Area
  9.3: Calculus and Parametric Equations
speed
  11.3: The Calculus of Motion
squeeze theorem
  1.3: Finding Limits Analytically
Sum Rule
  5.1: Antiderivatives and Indefinite Integration
surface area
  7.4: Arc Length and Surface Area
  9.3: Calculus and Parametric Equations
  9.5: Calculus and Polar Functions
  13.5: Surface Area
surface of revolution
  9.3: Calculus and Parametric Equations
  10.1: Introduction to Cartesian Coordinates in Space

T
tangent line
  12.7: Tangent Lines, Normal Lines, and Tangent Planes
tangent plane
  12.7: Tangent Lines, Normal Lines, and Tangent Planes
tangents (parametric equations)
  9.3: Calculus and Parametric Equations
Taylor polynomials
  8.7: Taylor Polynomials
Taylor series
  8.8: Taylor Series
telescoping series
  8.2: Infinite Series
the Product Rule
  2.4: The Product and Quotient Rules
total differential
  12.4: Differentiability and the Total Differential
trigonometric integrals
  6.3: Trigonometric Integrals
triple integral
  13.6: Volume Between Surfaces and Triple Integration

U
Unit Tangent Vector
  11.4: Unit Tangent and Normal Vectors

V
vector
  10.2: An Introduction to Vectors
vector function
  11.1: Vector–Valued Functions
vectors
  10: Vectors

velocity
11.3: The Calculus of Motion
Volume Between Surfaces
13.6: Volume Between Surfaces and Triple Integration
Volume by Shells
7.3: The Shell Method

W
washer method
7.2: Volume by Cross-Sectional Area- Disk and Washer Methods
work
7.5: Work

Glossary

absolute convergence | if the series $\sum_{n=1}^{\infty}|a_n|$ converges, the series $\sum_{n=1}^{\infty}a_n$ is said to converge absolutely

absolute error | if B is an estimate of some quantity having an actual value of A, then the absolute error is given by $|A-B|$

absolute extremum | if f has an absolute maximum or absolute minimum at c, we say f has an absolute extremum at c

absolute maximum | if $f(c) \geq f(x)$ for all x in the domain of f, we say f has an absolute maximum at c

absolute minimum | if $f(c) \leq f(x)$ for all x in the domain of f, we say f has an absolute minimum at c

absolute value function | $f(x) = \{-x, \text{ if } x < 0$\ifx,$x \geq 0$

acceleration | is the rate of change of the velocity, that is, the derivative of velocity

acceleration vector | the second derivative of the position vector

algebraic function | a function involving any combination of only the basic operations of addition, subtraction, multiplication, division, powers, and roots applied to an input variable x

alternating series | a series of the form $\sum_{n=1}^{\infty}(-1)^{n+1}b_n$ or $\sum_{n=1}^{\infty}(-1)^{n}b_n$, where $b_n \geq 0$, is called an alternating series

alternating series test | for an alternating series of either form, if $b_{n+1} \leq b_n$ for all integers $n \geq 1$ and $b_n \to 0$, then an alternating series converges

amount of change | the amount of a function $f(x)$ over an interval $[x,x+h] is f(x+h)-f(x)$

angular coordinate | θ the angle formed by a line segment connecting the origin to a point in the polar coordinate system with the positive radial (x) axis, measured counterclockwise

antiderivative | a function F such that $F'(x)=f(x)$ for all x in the domain of f is an antiderivative of f

arc length | the arc length of a curve can be thought of as the distance a person would travel along the path of the curve

arc-length function | a function $s(t)$ that describes the arc length of curve C as a function of t

arc-length parameterization | a reparameterization of a vector-valued function in which the parameter is equal to the arc length

arithmetic sequence | a sequence in which the difference between every pair of consecutive terms is the same is called an arithmetic sequence

asymptotically semi-stable solution | $y=k$ if it is neither asymptotically stable nor asymptotically unstable

asymptotically stable solution | $y=k$ if there exists $\varepsilon > 0$ such that for any value $c \in (k-\varepsilon, k+\varepsilon)$ the solution to the initial-value problem $y = f(x,y), y(x_0) = c$ approaches k as x approaches infinity

asymptotically unstable solution | $y=k$ if there exists $\varepsilon > 0$ such that for any value $c \in (k-\varepsilon, k+\varepsilon)$ the solution to the initial-value problem $y' = f(x,y), y(x_0) = c$ never approaches k as $x$approaches infinity

autonomous differential equation | an equation in which the right-hand side is a function of y alone

average rate of change | is a function $f(x)$ over an interval $[x,x+h]$ is $\frac{f(x+h)-f(a)}{b-a}$

average value of a function | (or f_{ave}) the average value of a function on an interval can be found by calculating the definite integral of the function and dividing that value by the length of the interval

average velocity | the change in an object's position divided by the length of a time period; the average velocity of an object over a time interval $[t,a]$ (if $t<a$ or $[a,t]$ if $t>a$), with a position given by $s(t)$, that is $v_{ave} = \dfrac{s(t)-s(a)}{t-a}$

base | the number b in the exponential function $f(x) = b^x$ and the logarithmic function $f(x) = \log_b x$

binomial series | the Maclaurin series for $f(x) = (1+x)^r$; it is given by $(1+x)^r = \sum_{n=0}^{\infty}\binom{r}{n}x^n = 1 + rx + \dfrac{r(r-1)}{2!}x^2 + \cdots + \dfrac{r(r-1)\cdots(r-n+1)}{n!}x^n + \cdots$ for $|x| < 1$

binormal vector | a unit vector orthogonal to the unit tangent vector and the unit normal vector

boundary conditions | the conditions that give the state of a system at different times, such as the position of a spring-mass system at two different times

boundary point | a point P_0 of R is a boundary point if every δ disk centered around P_0 contains points both inside and outside R

boundary-value problem | a differential equation with associated boundary conditions

bounded above | a sequence a_n is bounded above if there exists a constant M such that $a_n \leq M$ for all positive integers n

bounded below | a sequence a_n is bounded below if there exists a constant M such that $M \leq a_n$ for all positive integers n

bounded sequence | a sequence a_n is bounded if there exists a constant M such that $|a_n| \leq M$ for all positive integers n

cardioid | a plane curve traced by a point on the perimeter of a circle that is rolling around a fixed circle of the same radius; the equation of a cardioid is $r = a(1+\sin\theta)$ or $r = a(1+\cos\theta)$

carrying capacity | the maximum population of an organism that the environment can sustain indefinitely

catenary | a curve in the shape of the function $y = a \cdot \cosh(x/a)$ is a catenary; a cable of uniform density suspended between two supports assumes the shape of a catenary

center of mass | the point at which the total mass of the system could be concentrated without changing the moment

centroid | the centroid of a region is the geometric center of the region; laminas are often represented by regions in the plane; if the lamina has a constant density, the center of mass of the lamina depends only on the shape of the corresponding planar region; in this case, the center of mass of the lamina corresponds to the centroid of the representative region

chain rule | the chain rule defines the derivative of a composite function as the derivative of the outer function evaluated at the inner function times the derivative of the inner function

change of variables | the substitution of a variable, such as u, for an expression in the integrand

characteristic equation | the equation $a\lambda^2 + b\lambda + c = 0$ for the differential equation $ay''+by'+cy=0$

circulation | the tendency of a fluid to move in the direction of curve C. If C is a closed curve, then the circulation of $\vec{\mathbf{F}}$ along C is line integral $\int_C \vec{\mathbf{F}} \cdot \vec{\mathbf{T}}\, ds$, which we also denote $\oint_C \vec{\mathbf{F}} \cdot \vec{\mathbf{T}}\, ds$.

closed curve | a curve for which there exists a parameterization $\vec{\mathbf{r}}(t), a \leq t \leq b$, such that $\vec{\mathbf{r}}(a) = \vec{\mathbf{r}}(b)$, and the curve is traversed exactly once

closed curve | a curve that begins and ends at the same point

closed set | a set S that contains all its boundary points

comparison test | If $0 \leq a_n \leq b_n$ for all $n \geq N$ and $\sum_{n=1}^{\infty} b_n$ converges, then $\sum_{n=1}^{\infty} a_n$ converges; if $a_n \geq b_n \geq 0$ for all $n \geq N$ and $\sum_{n=1}^{\infty} b_n$ diverges, then $\sum_{n=1}^{\infty} a_n$ diverges.

complementary equation | for the nonhomogeneous linear differential equation

$$a+2(x)y\\ ''+a_1(x)y\\ '+a_0(x)y\\ =r(x),$$

the associated homogeneous equation, called the *complementary equation*, is

$$a_2(x)y''\\ +a_1(x)y\\ '+a_0(x)y=0$$

component | a scalar that describes either the vertical or horizontal direction of a vector

component functions | the component functions of the vector-valued function $\vec{\mathbf{r}}(t) = f(t)\hat{\mathbf{i}} + g(t)\hat{\mathbf{j}}$ are $f(t)$ and $g(t)$, and the component functions of the vector-valued function $\vec{\mathbf{r}}(t) = f(t)\hat{\mathbf{i}} + g(t)\hat{\mathbf{j}} + h(t)\hat{\mathbf{k}}$ are $f(t)$, $g(t)$ and $h(t)$

composite function | given two functions f and g, a new function, denoted $g \circ f$, such that $(g \circ f)(x) = g(f(x))$

computer algebra system (CAS) | technology used to perform many mathematical tasks, including integration

concave down | if f is differentiable over an interval I and f' is decreasing over I, then f is concave down over I

concave up | if f is differentiable over an interval I and f' is increasing over I, then f is concave up over I

concavity | the upward or downward curve of the graph of a function

concavity test | suppose f is twice differentiable over an interval I; if $f''>0$ over I, then f is concave up over I; if $f''<$ over I, then f is concave down over I

conditional convergence | if the series $\sum_{n=1}^{\infty} a_n$ converges, but the series $\sum_{n=1}^{\infty} |a_n|$ diverges, the series $\sum_{n=1}^{\infty} a_n$ is said to converge conditionally

conic section | a conic section is any curve formed by the intersection of a plane with a cone of two nappes

connected region | a region in which any two points can be connected by a path with a trace contained entirely inside the region

connected set | an open set S that cannot be represented as the union of two or more disjoint, nonempty open subsets

conservative field | a vector field for which there exists a scalar function f such that $\vec{\nabla} f = \vec{\mathbf{F}}$

constant multiple law for limits | the limit law

$$\lim_{x \to a} cf(x) \\ = c \cdot \lim_{x \to a} f(x) \\ = cL$$

constant multiple rule | the derivative of a constant c multiplied by a function f is the same as the constant multiplied by the derivative: $\frac{d}{dx}\big(cf(x)\big) = cf'(x)$

constant rule | the derivative of a constant function is zero: $\frac{d}{dx}(c) = 0$, where c is a constant

constraint | an inequality or equation involving one or more variables that is used in an optimization problem; the constraint enforces a limit on the possible solutions for the problem

continuity at a point | A function $f(x)$ is continuous at a point a if and only if the following three conditions are satisfied: (1) $f(a)$ is defined, (2) $\lim_{x \to a} f(x)$ exists, and (3) $\lim x \to af(x) = f(a)$

continuity from the left | A function is continuous from the left at b if $\lim_{x \to b^-} f(x) = f(b)$

continuity from the right | A function is continuous from the right at a if $\lim_{x \to a^+} f(x) = f(a)$

continuity over an interval | a function that can be traced with a pencil without lifting the pencil; a function is continuous over an open interval if it is continuous at every point in the interval; a function $f(x)$ is continuous over a closed interval of the form $[a, b]$ if it is continuous at every point in (a, b), and it is continuous from the right at a and from the left at b

contour map | a plot of the various level curves of a given function $f(x, y)$

convergence of a series | a series converges if the sequence of partial sums for that series converges

convergent sequence | a convergent sequence is a sequence a_n for which there exists a real number L such that a_n is arbitrarily close to L as long as n is sufficiently large

coordinate plane | a plane containing two of the three coordinate axes in the three-dimensional coordinate system, named by the axes it contains: the xy-plane, xz-plane, or the yz-plane

critical point | if $f'(c) = 0$ or $f'(c)$ is undefined, we say that c is a critical point of f

critical point of a function of two variables | the point (x_0, y_0) is called a critical point of $f(x, y)$ if one of the two following conditions holds: 1. $f_x(x_0, y_0) = f_y(x_0, y_0) = 0$ 2. At least one of $f_x(x_0, y_0)$ and $f_y(x_0, y_0)$ do not exist

cross product | $\vec{\mathbf{u}} \times \vec{\mathbf{v}} = (u_2v_3 - u_3v_2)\hat{\mathbf{i}} - (u_1v_3 - u_3v_1)\hat{\mathbf{j}} + (u_1v_2 - u_2v_1)\hat{\mathbf{k}}$, where $\vec{\mathbf{u}} = \langle u_1, u_2, u_3 \rangle$ and $\vec{\mathbf{v}} = \langle v_1, v_2, v_3 \rangle$ determinant a real number associated with a square matrix parallelepiped a three-dimensional prism with six faces that are parallelograms torque the effect of a force that causes an object to rotate triple scalar product the dot product of a vector with the cross product of two other vectors: $\vec{\mathbf{u}} \cdot (\vec{\mathbf{v}} \times \vec{\mathbf{w}})$ vector product the cross product of two vectors.

cross-section | the intersection of a plane and a solid object

cubic function | a polynomial of degree 3; that is, a function of the form $f(x) = ax^3 + bx^2 + cx + d$, where $a \neq 0$

curl | the curl of vector field $\vec{\mathbf{F}} = \langle P, Q, R \rangle$, denoted $\vec{\nabla} \times \vec{\mathbf{F}}$ is the “determinant” of the matrix

$$\left| \hat{\mathbf{i}} \quad \hat{\mathbf{j}} \quad \hat{\mathbf{k}} \; \frac{\partial}{\partial x} \quad \frac{\partial}{\partial y} \quad \frac{\partial}{\partial z} \; P \quad Q \quad R \right|$$

.

and is given by the expression $(R_y - Q_z)\,\hat{\mathbf{i}} + (P_z - R_x)\,\hat{\mathbf{j}} + (Q_x - P_y)\,\hat{\mathrm{k}}$; it measures the tendency of particles at a point to rotate about the axis that points in the direction of the curl at the point

curvature | the derivative of the unit tangent vector with respect to the arc-length parameter

cusp | a pointed end or part where two curves meet

cycloid | the curve traced by a point on the rim of a circular wheel as the wheel rolls along a straight line without slippage

cylinder | a set of lines parallel to a given line passing through a given curve

cylindrical coordinate system | a way to describe a location in space with an ordered triple (r, θ, z), where (r, θ) represents the polar coordinates of the point’s projection in the xy-plane, and z represents the point’s projection onto the z-axis

decreasing on the interval

definite integral | a primary operation of calculus; the area between the curve and the x-axis over a given interval is a definite integral

definite integral of a vector-valued function | the vector obtained by calculating the definite integral of each of the component functions of a given vector-valued function, then using the results as the components of the resulting function

degree | for a polynomial function, the value of the largest exponent of any term

density function | a density function describes how mass is distributed throughout an object; it can be a linear density, expressed in terms of mass per unit length; an area density, expressed in terms of mass per unit area; or a volume density, expressed in terms of mass per unit volume; weight-density is also used to describe weight (rather than mass) per unit volume

dependent variable | the output variable for a function

derivative | the slope of the tangent line to a function at a point, calculated by taking the limit of the difference quotient, is the derivative

derivative function | gives the derivative of a function at each point in the domain of the original function for which the derivative is defined

derivative of a vector-valued function | the derivative of a vector-valued function $\vec{\mathbf{r}}(t)$ is $\vec{\mathbf{r}}'(t) = \lim_{\Delta t \to 0} \frac{\vec{\mathbf{r}}(t+\Delta t) \;\; \vec{\mathbf{r}}(t)}{\Delta t}$, provided the limit exists

difference law for limits | the limit law

$$\lim_{x \to a}(f(x) \\ - g(x)) \\ = \lim_{x \to a} f(x) \\ - \lim_{x \to a} g(x) \\ = L - M$$

difference quotient | of a function $f(x)$ at a is given by $\frac{f(a+h) - f(a)}{h}$ or $\frac{f(x) - f(a)}{x - a}$

difference rule | the derivative of the difference of a function f and a function g is the same as the difference of the derivative of f and the derivative of g: $\frac{d}{dx}\big(f(x) - g(x)\big) = f'(x) - g'(x)$

differentiable | a function $f(x, y)$ is differentiable at (x_0, y_0) if $f(x, y)$ can be expressed in the form $f(x, y) = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) + E(x, y)$, where the error term $E(x, y)$ satisfies $\lim_{(x,y) \to (x_0,y_0)} \frac{E(x, y)}{\sqrt{(x - x_0)^2 + (y - y_0)^2}} = 0$

differentiable at

differentiable function | a function for which $f'(x)$ exists is a differentiable function

differentiable on

differential | the differential dx is an independent variable that can be assigned any nonzero real number; the differential dy is defined to be $dy = f'(x)\,dx$

differential calculus | the field of calculus concerned with the study of derivatives and their applications

differential equation | an equation involving a function $y = y(x)$ and one or more of its derivatives

differential form | given a differentiable function $y = f'(x)$, the equation $dy = f'(x)\,dx$ is the differential form of the derivative of y with respect to x

differentiation | the process of taking a derivative

direction angles | the angles formed by a nonzero vector and the coordinate axes

direction cosines | the cosines of the angles formed by a nonzero vector and the coordinate axes

direction field (slope field) | a mathematical object used to graphically represent solutions to a first-order differential equation; at each point in a direction field, a line segment appears whose slope is equal to the slope of a solution to the differential equation passing through that point

direction vector | a vector parallel to a line that is used to describe the direction, or orientation, of the line in space

directional derivative | the derivative of a function in the direction of a given unit vector

directrix | a directrix (plural: directrices) is a line used to construct and define a conic section; a parabola has one directrix; ellipses and hyperbolas have two

discontinuity at a point | A function is discontinuous at a point or has a discontinuity at a point if it is not continuous at the point

discriminant | the value $4AC - B^2$, which is used to identify a conic when the equation contains a term involving xy, is called a discriminant

discriminant | the discriminant of the function $f(x, y)$ is given by the formula $D = f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - (f_{xy}(x_0, y_0))^2$

disk method | a special case of the slicing method used with solids of revolution when the slices are disks

divergence | the divergence of a vector field $\vec{\mathbf{F}} = \langle P,Q,R\rangle$, denoted $\vec{\nabla} \times \vec{\mathbf{F}}$, is $P_x + Q_y + R_z$; it measures the “outflowing-ness” of a vector field

divergence of a series | a series diverges if the sequence of partial sums for that series diverges

divergence test | if $\lim\limits_{n\to\infty} a_n \neq 0$, then the series $\sum\limits_{n=1}^{\infty} a_n$ diverges

divergent sequence | a sequence that is not convergent is divergent

domain | the set of inputs for a function

dot product or scalar product | $\vec{\mathbf{u}} \cdot \vec{\mathbf{v}} = u_1v_1 + u_2v_2 + u_3v_3$ where $\vec{\mathbf{u}} = \langle u_1,u_2,u_3\rangle$ and $\vec{\mathbf{v}} = \langle v_1,v_2,v_3\rangle$

double integral | of the function $f(x,y)$ over the region R in the xy-plane is defined as the limit of a double Riemann sum,

$$\iint_R f(x,y)\,dA = \lim_{m,n\to\infty} \sum_{i=1}^{m} \sum_{j=1}^{n} f(x^*_{ij},y^*_{ij}) \Delta A.$$

double Riemann sum | of the function $f(x,y)$ over a rectangular region R is

$$\sum_{i=1}^{m} \sum_{j=1}^{n} f(x^*_{ij},y^*_{ij}) \Delta A,$$

where R is divided into smaller subrectangles R_{ij} and (x^*_{ij},y^*_{ij}) is an arbitrary point in R_{ij}

doubling time | if a quantity grows exponentially, the doubling time is the amount of time it takes the quantity to double, and is given by $(\ln 2)/k$

eccentricity | the eccentricity is defined as the distance from any point on the conic section to its focus divided by the perpendicular distance from that point to the nearest directrix

ellipsoid | a three-dimensional surface described by an equation of the form $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$; all traces of this surface are ellipses

elliptic cone | a three-dimensional surface described by an equation of the form $\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0$; traces of this surface include ellipses and intersecting lines

elliptic paraboloid | a three-dimensional surface described by an equation of the form $z = \frac{x^2}{a^2} + \frac{y^2}{b^2}$; traces of this surface include ellipses and parabolas

end behavior | the behavior of a function as $x \to \infty$ and $x \to -\infty$

epsilon-delta definition of the limit | $\lim\limits_{x\to a} f(x) = L$ if for every $\varepsilon > 0$, there exists a $\delta > 0$ such that if $0 < |x - a| < \delta$, then $|f(x) - L| < \varepsilon$

equilibrium solution | any solution to the differential equation of the form $y = c$, where c is a constant

equivalent vectors | vectors that have the same magnitude and the same direction

Euler’s Method | a numerical technique used to approximate solutions to an initial-value problem

even function | a function is even if $f(-x) = f(x)$ for all x in the domain of f

explicit formula | a sequence may be defined by an explicit formula such that $a_n = f(n)$

exponent | the value x in the expression b^x

exponential decay | systems that exhibit exponential decay follow a model of the form $y = y_0e^{-kt}$

exponential growth | systems that exhibit exponential growth follow a model of the form $y = y_0e^{kt}$

extreme value theorem | if f is a continuous function over a finite, closed interval, then f has an absolute maximum and an absolute minimum

Fermat’s theorem | if f has a local extremum at c, then c is a critical point of f

first derivative test | let f be a continuous function over an interval I containing a critical point c such that f is differentiable over I except possibly at c; if f' changes sign from positive to negative as x increases through c, then f has a local maximum at c; if f' changes sign from negative to positive as x increases through c, then f has a local minimum at c; if f' does not change sign as x increases through c, then f does not have a local extremum at c

flux | the rate of a fluid flowing across a curve in a vector field; the flux of vector field $\vec{\mathbf{F}}$ across plane curve C is line integral $\int_C \vec{\mathbf{F}} \cdot \frac{\vec{\mathbf{n}}(t)}{\|\vec{\mathbf{n}}(t)\|}\,ds$

flux integral | another name for a surface integral of a vector field; the preferred term in physics and engineering

focal parameter | the focal parameter is the distance from a focus of a conic section to the nearest directrix

focus | a focus (plural: foci) is a point used to construct and define a conic section; a parabola has one focus; an ellipse and a hyperbola have two

formal definition of an infinite limit | $\lim\limits_{x\to a} f(x) = \infty$ if for every $M > 0$, there exists a $\delta > 0$ such that if $0 < |x - a| < \delta$, then $f(x) > M$ $\lim\limits_{x\to a} f(x) = -\infty$ if for every $M > 0$, there exists a $\delta > 0$ such that if $0 < |x - a| < \delta$, then $f(x) < -M$

Frenet frame of reference | (TNB frame) a frame of reference in three-dimensional space formed by the unit tangent vector, the unit normal vector, and the binormal vector

frustum | a portion of a cone; a frustum is constructed by cutting the cone with a plane parallel to the base

Fubini’s theorem | if $f(x,y)$ is a function of two variables that is continuous over a rectangular region $R = \{(x,y) \in \mathbb{R}^2 \mid a \le x \le b,\ c \le y \le d\}$, then the double integral of f over the region equals an iterated integral,

$$\iint_R f(x,y)\,dA = \int_a^b \int_c^d f(x,y)\,dx\,dy = \int_c^d \int_a^b f(x,y)\,dx\,dy$$

function | a set of inputs, a set of outputs, and a rule for mapping each input to exactly one output

function of two variables | a function $z = f(x,y)$ that maps each ordered pair (x,y) in a subset D of R^2 to a unique real number z

Fundamental Theorem for Line Integrals | the value of line integral $\int_C \vec{\nabla} f \cdot d\vec{\mathbf{r}}$ depends only on the value of f at the endpoints of C : $\int_C \vec{\nabla} f \cdot d\vec{\mathbf{r}} = f(\vec{\mathbf{r}}(b)) - f(\vec{\mathbf{r}}(a))$

fundamental theorem of calculus | (also, evaluation theorem) we can evaluate a definite integral by evaluating the antiderivative of the integrand at the endpoints of the interval and subtracting

fundamental theorem of calculus | uses a definite integral to define an antiderivative of a function

fundamental theorem of calculus | the theorem, central to the entire development of calculus, that establishes the relationship between differentiation and integration

general form | an equation of a conic section written as a general second-degree equation

general form of the equation of a plane | an equation in the form $ax + by + cz + d = 0$, where $\vec{\mathbf{n}} = \langle a,b,c\rangle$ is a normal vector of the plane, $P = (x_0,y_0,z_0)$ is a point on the plane, and $d = -ax_0 - by_0 - cz_0$

general solution (or family of solutions) | the entire set of solutions to a given differential equation

generalized chain rule | the chain rule extended to functions of more than one independent variable, in which each independent variable may depend on one or more other variables

geometric sequence | a sequence a_n in which the ratio a_{n+1}/a_n is the same for all positive integers n is called a geometric sequence

geometric series | a geometric series is a series that can be written in the form $\sum\limits_{n=1}^{\infty} ar^{n-1} = a + ar + ar^2 + ar^3 + \cdots$

gradient | the gradient of the function $f(x,y)$ is defined to be $\vec{\nabla} f(x,y) = (\partial f/\partial x)\,\hat{\mathbf{i}} + (\partial f/\partial y)\,\hat{\mathbf{j}}$, which can be generalized to a function of any number of independent variables

gradient field | a vector field $\vec{\mathbf{F}}$ for which there exists a scalar function f such that $\vec{\nabla} f = \vec{\mathbf{F}}$; in other words, a vector field that is the gradient of a function; such vector fields are also called *conservative*

graph of a function | the set of points (x,y) such that x is in the domain of f and $y = f(x)$

graph of a function of two variables | a set of ordered triples (x,y,z) that satisfies the equation $z = f(x,y)$ plotted in three-dimensional Cartesian space

Green’s theorem | relates the integral over a connected region to an integral over the boundary of the region

grid curves | curves on a surface that are parallel to grid lines in a coordinate plane

growth rate | the constant $r > 0$ in the exponential growth function $P(t) = P_0e^{rt}$

half-life | if a quantity decays exponentially, the half-life is the amount of time it takes the quantity to be reduced by half. It is given by $(\ln 2)/k$

harmonic series | the harmonic series takes the form $\sum\limits_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} +$

heat flow | a vector field proportional to the negative temperature gradient in an object

helix | a three-dimensional curve in the shape of a spiral

higher-order derivative | a derivative of a derivative, from the second derivative to the n^{th} derivative, is called a higher-order derivative

higher-order partial derivatives | second-order or higher partial derivatives, regardless of whether they are mixed partial derivatives

homogeneous linear equation | a second-order differential equation that can be written in the form $a_2(x)y''+a_1(x)y'+a_0(x)y=r(x)$, but $r(x)=0$ for every value of x

Hooke's law | this law states that the force required to compress (or elongate) a spring is proportional to the distance the spring has been compressed (or stretched) from equilibrium; in other words, $F=kx$, where k is a constant

horizontal asymptote | if $\lim\limits_{x\to\infty} f(x) = L$ or $\lim\limits_{x\to-\infty} f(x)=L$, then $y=L$ is a horizontal asymptote of f

horizontal line test | a function f is one-to-one if and only if every horizontal line intersects the graph of f, at most, once

hydrostatic pressure | the pressure exerted by water on a submerged object

hyperbolic functions | the functions denoted $\sinh$, $\cosh$, $\tanh$, csch, sech, and $\coth$, which involve certain combinations of e^x and e^{-x}

hyperboloid of one sheet | a three-dimensional surface described by an equation of the form $\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=1$; traces of this surface include ellipses and hyperbolas

hyperboloid of two sheets | a three-dimensional surface described by an equation of the form $\frac{z^2}{c^2}-\frac{x^2}{a^2}-\frac{y^2}{b^2}=1$; traces of this surface include ellipses and hyperbolas

implicit differentiation | is a technique for computing $\frac{dy}{dx}$ for a function defined by an equation, accomplished by differentiating both sides of the equation (remembering to treat the variable y as a function) and solving for $\frac{dy}{dx}$

improper double integral | a double integral over an unbounded region or of an unbounded function

improper integral | an integral over an infinite interval or an integral of a function containing an infinite discontinuity on the interval; an improper integral is defined in terms of a limit. The improper integral converges if this limit is a finite real number; otherwise, the improper integral diverges

increasing on the interval

indefinite integral | the most general antiderivative of $f(x)$ is the indefinite integral of f; we use the notation $\int f(x)\,dx$ to denote the indefinite integral of f

indefinite integral of a vector-valued function | a vector-valued function with a derivative that is equal to a given vector-valued function

independence of path | a vector field $\vec{\mathbf{F}}$ has path independence if $\int_{C_1} \vec{\mathbf{F}}\cdot d\vec{\mathbf{r}} = \int_{C_2} \vec{\mathbf{F}}\cdot d\vec{\mathbf{r}}$ for any curves C_1 and C_2 in the domain of $\vec{\mathbf{F}}$ with the same initial points and terminal points

independent variable | the input variable for a function

indeterminate forms | When evaluating a limit, the forms $\frac{0}{0}$,$\infty/\infty, 0\cdot\infty, \infty-\infty, 0^0, \infty^0$, and 1^∞ are considered indeterminate because further analysis is required to determine whether the limit exists and, if so, what its value is.

index variable | the subscript used to define the terms in a sequence is called the index

infinite discontinuity | An infinite discontinuity occurs at a point a if $\lim\limits_{x\to a^-} f(x) = \pm\infty$ or $\lim\limits_{x\to a^+} f(x) = \pm\infty$

infinite limit | A function has an infinite limit at a point a if it either increases or decreases without bound as it approaches a

infinite limit at infinity | a function that becomes arbitrarily large as x becomes large

infinite series | an infinite series is an expression of the form $a_1+a_2+a_3+\cdots=\sum\limits_{n=1}^{\infty} a_n$

inflection point | if f is continuous at c and f changes concavity at c, the point $(c,f(c))$ is an inflection point of f

initial point | the starting point of a vector

initial population | the population at time $t=0$

initial value problem | a problem that requires finding a function y that satisfies the differential equation $\frac{dy}{dx}=f(x)$ together with the initial condition $y(x_0)=y_0$

initial value(s) | a value or set of values that a solution of a differential equation satisfies for a fixed value of the independent variable

initial velocity | the velocity at time $t=0$

initial-value problem | a differential equation together with an initial value or values

instantaneous rate of change | the rate of change of a function at any point along the function a, also called $f'(a)$, or the derivative of the function at a

instantaneous velocity | The instantaneous velocity of an object with a position function that is given by $s(t)$ is the value that the average velocities on intervals of the form $[t,a]$ and $[a,t]$ approach as the values of t move closer to a, provided such a value exists

integrable function | a function is integrable if the limit defining the integral exists; in other words, if the limit of the Riemann sums as n goes to infinity exists

integral calculus | the study of integrals and their applications

integral test | for a series $\sum\limits_{n=1}^{\infty} a_n$ with positive terms a_n, if there exists a continuous, decreasing function f such that $f(n)=a_n$ for all positive integers n, then

$$\sum_{n=1}^{\infty} a_n$$

and

$$\int_1^{\infty} f(x)\,dx$$

either both converge or both diverge

integrand | the function to the right of the integration symbol; the integrand includes the function being integrated

integrating factor | any function $f(x)$ that is multiplied on both sides of a differential equation to make the side involving the unknown function equal to the derivative of a product of two functions

integration by parts | a technique of integration that allows the exchange of one integral for another using the formula $\int u\,dv=uv-\int v\,du$

integration by substitution | a technique for integration that allows integration of functions that are the result of a chain-rule derivative

integration table | a table that lists integration formulas

interior point | a point P_0 of $\mathbb{R}$ is a boundary point if there is a δ disk centered around P_0 contained completely in $\mathbb{R}$

Intermediate Value Theorem | Let f be continuous over a closed bounded interval $[a,b]$ if z is any real number between $f(a)$ and $f(b)$, then there is a number c in $[a,b]$ satisfying $f(c)=z$

intermediate variable | given a composition of functions (e.g., $f(x(t),y(t))$), the intermediate variables are the variables that are independent in the outer function but dependent on other variables as well; in the function $f(x(t),y(t))$, the variables x and y are examples of intermediate variables

interval of convergence | the set of real numbers x for which a power series converges

intuitive definition of the limit | If all values of the function $f(x)$ approach the real number L as the values of $x(\neq a)$ approach a, $f(x)$ approaches L

inverse function | for a function f, the inverse function f^{-1} satisfies $f^{-1}(y)=x$ if $f(x)=y$

inverse hyperbolic functions | the inverses of the hyperbolic functions where $\cosh$ and sech are restricted to the domain $[0,\infty)$;each of these functions can be expressed in terms of a composition of the natural logarithm function and an algebraic function

inverse trigonometric functions | the inverses of the trigonometric functions are defined on restricted domains where they are one-to-one functions

iterated integral | for a function $f(x,y)$ over the region R is a. $\int_a^b\int_c^d f(x,y)\,dx\,dy = \int_a^b \left[\int_c^d f(x,y)\,dy\right]dx$, b. $\int_c^d\int_a^b f(x,y)\;dx\;dy = \int_c^d\left[\int_a^b f(x,y)\,dx\right]\;dy$, where a,b,c, and d are any real numbers and $R=[a,b]\times[c,d]$

iterative process | process in which a list of numbers $x_0,x_1,x_2,x_3\ldots$ is generated by starting with a number x_0 and defining $x_n=F(x_{n-1})$ for $n\geq 1$

Jacobian | the Jacobian $J(u,v)$ in two variables is a 2×2 determinant:

$$J(u,v)= \left|\begin{array}{cc} \frac{\partial x}{\partial u}\frac{\partial y}{\partial u} & \frac{\partial x}{\partial v}\frac{\partial y}{\partial v}\end{array}\right|;$$

the Jacobian $J(u,v,w)$ in three variables is a 3×3 determinant:

$$J(u,v,w)= \left|\begin{array}{cc} \frac{\partial x}{\partial u}\frac{\partial y}{\partial u}\frac{\partial z}{\partial u} & \frac{\partial x}{\partial v}\frac{\partial y}{\partial v} \\ \frac{\partial z}{\partial v} & \frac{\partial x}{\partial w}\frac{\partial y}{\partial w}\frac{\partial z}{\partial w}\end{array}\right|$$

jump discontinuity | A jump discontinuity occurs at a point a if $\lim\limits_{x\to a^-} f(x)$ and $\lim\limits_{x\to a^+} f(x)$ both exist, but $\lim\limits_{x\to a^-} f(x)\neq\lim\limits_{x\to a^+} f(x)$

Kepler’s laws of planetary motion | three laws governing the motion of planets, asteroids, and comets in orbit around the Sun

Lagrange multiplier | the constant (or constants) used in the method of Lagrange multipliers; in the case of one constant, it is represented by the variable λ

lamina | a thin sheet of material; laminas are thin enough that, for mathematical purposes, they can be treated as if they are two-dimensional

left-endpoint approximation | an approximation of the area under a curve computed by using the left endpoint of each subinterval to calculate the height of the vertical sides of each rectangle

level curve of a function of two variables | the set of points satisfying the equation $f(x,y)=c$ for some real number c in the range of f

level surface of a function of three variables | the set of points satisfying the equation $f(x,y,z)=c$ for some real number c in the range of f

limaçon | the graph of the equation $r=a+b\sin\theta$ or $r=a+b\cos\theta$. If $a=b$ then the graph is a cardioid

limit | the process of letting x or t approach a in an expression; the limit of a function $f(x)$ as x approaches a is the value that $f(x)$ approaches as x approaches a

limit at infinity | a function that approaches a limit value L as x becomes large

limit comparison test | Suppose $a_n, b_n \geq 0$ for all $n \geq 1$. If $\lim\limits_{n\to\infty} a_n/b_n \to L \neq 0$, then $\sum\limits_{n=1}^{\infty} a_n$ and $\sum\limits_{n=1}^{\infty} b_n$ both converge or both diverge; if $\lim\limits_{n\to\infty} a_n/b_n \to 0$ and $\sum\limits_{n=1}^{\infty} b_n$ converges, then $\sum\limits_{n=1}^{\infty} a_n$ converges. If $\lim\limits_{n\to\infty} a_n/b_n \to \infty$, and $\sum\limits_{n=1}^{\infty} b_n$ diverges, then $\sum\limits_{n=1}^{\infty} a_n$ diverges.

limit laws | the individual properties of limits; for each of the individual laws, let $f(x)$ and $g(x)$ be defined for all $x \neq a$ over some open interval containing a; assume that L and M are real numbers so that $\lim_{x\to a} f(x)=L$ and $\lim_{x\to a} g(x)=M$; let c be a constant

limit of a sequence | the real number LL to which a sequence converges is called the limit of the sequence

limit of a vector-valued function | a vector-valued function $\vec{\mathbf{r}}(t)$ has a limit $\vec{\mathbf{L}}$ as t approaches a if $\lim t \to a \left|\vec{\mathbf{r}}(t)-\vec{\mathbf{L}}\right| = 0$

limits of integration | these values appear near the top and bottom of the integral sign and define the interval over which the function should be integrated

line integral | the integral of a function along a curve in a plane or in space

linear | description of a first-order differential equation that can be written in the form $a(x)y'+b(x)y=c(x)$

linear approximation | the linear function $L(x)=f(a)+f'(a)(x-a)$ is the linear approximation of f at $x=a$

linear approximation | given a function $f(x,y)$ and a tangent plane to the function at a point (x_0,y_0), we can approximate $f(x,y)$ for points near (x_0,y_0) using the tangent plane formula

linear function | a function that can be written in the form $f(x)=mx+b$

linearly dependent | a set of functions $f_1(x), f_2(x), \ldots, f_n(x)$ for whichthere are constants $c_1, c_2, \ldots c_n$, not all zero, such that $c_1f_1(x)+c_2f_2(x)+\cdots+c_nf_n(x)=0$ for all \(x\) in the interval of interest

linearly independent | a set of functions $f_1(x), f_2(x), \ldots, f_n(x)$ for which there are no constants $c_1, c_2, \ldots c_n$, such that $c_1f_1(x)+c_2f_2(x)+\cdots+c_nf_n(x)=0$ for all \(x\) in the interval of interest

local extremum | if f has a local maximum or local minimum at c, we say f has a local extremum at c

local maximum | if there exists an interval I such that $f(c) \geq f(x)$ for all $x \in I$, we say f has a local maximum at c

local minimum | if there exists an interval I such that $f(c) \leq f(x)$ for all $x \in I$, we say f has a local minimum at c

logarithmic differentiation | is a technique that allows us to differentiate a function by first taking the natural logarithm of both sides of an equation, applying properties of logarithms to simplify the equation, and differentiating implicitly

logarithmic function | a function of the form $f(x)=\log_b(x)$ for some base $b>0$, $b\neq 1$ such that $y=\log_b(x)$ if and only if $b^y=x$

logistic differential equation | a differential equation that incorporates the carrying capacity K and growth rate rr into a population model

lower sum | a sum obtained by using the minimum value of $f(x)$ on each subinterval

L’Hôpital’s rule | If f and g are differentiable functions over an interval a, except possibly at a, and $\lim\limits_{x\to a} f(x)=0=\lim\limits_{x\to a} g(x)$ or $\lim\limits_{x\to a} f(x)$ and $\lim\limits_{x\to a} g(x)$ are infinite, then $\lim\limits_{x\to a}\dfrac{f(x)}{g(x)}=\lim\limits_{x\to a}\dfrac{f'(x)}{g'(x)}$, assuming the limit on the right exists or is ∞ or $-\infty$.

Maclaurin polynomial | a Taylor polynomial centered at 0; the n^{th}-degree Taylor polynomial for f at 0 is the n^{th}-degree Maclaurin polynomial for f

Maclaurin series | a Taylor series for a function f at $x=0$ is known as a Maclaurin series for f

magnitude | the length of a vector

major axis | the major axis of a conic section passes through the vertex in the case of a parabola or through the two vertices in the case of an ellipse or hyperbola; it is also an axis of symmetry of the conic; also called the transverse axis

marginal cost | is the derivative of the cost function, or the approximate cost of producing one more item

marginal profit | is the derivative of the profit function, or the approximate profit obtained by producing and selling one more item

marginal revenue | is the derivative of the revenue function, or the approximate revenue obtained by selling one more item

mass flux | the rate of mass flow of a fluid per unit area, measured in mass per unit time per unit area

mathematical model | A method of simulating real-life situations with mathematical equations

mean value theorem | if f is continuous over $[a,b]$ and differentiable over (a,b), then there exists $c\in(a,b)$ such that $f'(c)=\frac{f(b)-f(a)}{b-a}$

mean value theorem for integrals | guarantees that a point c exists such that $f(c)$ is equal to the average value of the function

method of cylindrical shells | a method of calculating the volume of a solid of revolution by dividing the solid into nested cylindrical shells; this method is different from the methods of disks or washers in that we integrate with respect to the opposite variable

method of Lagrange multipliers | a method of solving an optimization problem subject to one or more constraints

method of undetermined coefficients | a method that involves making a guess about the form of the particular solution, then solving for the coefficients in the guess

method of variation of parameters | a method that involves looking for particular solutions in the form $y_p(x)=u(x)y_1(x)+v(x)y_2(x)$, where y_1 and y_2 are linearly independent solutions to the complementary equations, and then solving a system of equations to find $u(x)$ and $v(x)$

midpoint rule | a rule that uses a Riemann sum of the form $M_n=\sum\limits_{i=1}^{n} f(m_i)\Delta x$, where m_i is the midpoint of the i^{th} subinterval to approximate $\int_a^b f(x)\,dx$

minor axis | the minor axis is perpendicular to the major axis and intersects the major axis at the center of the conic, or at the vertex in the case of the parabola; also called the conjugate axis

mixed partial derivatives | second-order or higher partial derivatives, in which at least two of the differentiations are with respect to different variables

moment | if n masses are arranged on a number line, the moment of the system with respect to the origin is given by $M=\sum\limits_{i=1}^{n} m_ix_i$; if, instead, we consider a region in the plane, bounded above by a function $f(x)$ over an interval $[a,b]$, then the moments of the region with respect to the x- and y-axes are given by $M_x=\rho\int_a^b \frac{[f(x)]^2}{2}\,dx$ and $M_y=\rho\int_a^b xf(x)\,dx$, respectively

monotone sequence | an increasing or decreasing sequence

multivariable calculus | the study of the calculus of functions of two or more variables

nappe | a nappe is one half of a double cone

natural exponential function | the function $f(x)=e^x$

natural logarithm | the function $\ln x=\log_e x$

net change theorem | if we know the rate of change of a quantity, the net change theorem says the future quantity is equal to the initial quantity plus the integral of the rate of change of the quantity

net signed area | the area between a function and the x-axis such that the area below the x-axis is subtracted from the area above the x-axis; the result is the same as the definite integral of the function

Newton’s method | method for approximating roots of $f(x)=0$; using an initial guess x_0; each subsequent approximation is defined by the equation $x_n=x_{n-1}-\frac{f(x_{n-1})}{f'(x_{n-1})}$

nonelementary integral | an integral for which the antiderivative of the integrand cannot be expressed as an elementary function

nonhomogeneous linear equation | a second-order differential equation that can be written in the form $a_2(x)y''+a_1(x)y'+a_0(x)y=r(x)$, but $r(x)\neq 0$ for some value of x

normal component of acceleration | the coefficient of the unit normal vector $\vec{\mathbf{N}}$ when the acceleration vector is written as a linear combination of $\vec{\mathbf{T}}$ and $\vec{\mathbf{N}}$

normal plane | a plane that is perpendicular to a curve at any point on the curve

normal vector | a vector perpendicular to a plane

normalization | using scalar multiplication to find a unit vector with a given direction

number e | as m gets larger, the quantity $(1 + (1/m)^m$ gets closer to some real number; we define that real number to be e; the value of e is approximately 2.718282

numerical integration | the variety of numerical methods used to estimate the value of a definite integral, including the midpoint rule, trapezoidal rule, and Simpson's rule

objective function | the function that is to be maximized or minimized in an optimization problem

oblique asymptote | the line $y = mx + b$ if $f(x)$ approaches it as $x \to \infty$ or$x \to -\infty$

octants | the eight regions of space created by the coordinate planes

odd function | a function is odd if $f(-x) = -f(x)$ for all x in the domain of f

one-sided limit | A one-sided limit of a function is a limit taken from either the left or the right

one-to-one function | a function f is one-to-one if $f(x_1) \neq f(x_2)$ if $x_1 \neq x_2$

one-to-one transformation | a transformation $T : G \to R$ defined as $T(u, v) = (x, y)$ is said to be one-to-one if no two points map to the same image point

open set | a set S that contains none of its boundary points

optimization problem | calculation of a maximum or minimum value of a function of several variables, often using Lagrange multipliers

optimization problems | problems that are solved by finding the maximum or minimum value of a function

order of a differential equation | the highest order of any derivative of the unknown function that appears in the equation

orientation | the direction that a point moves on a graph as the parameter increases

orientation of a curve | the orientation of a curve C is a specified direction of C

orientation of a surface | if a surface has an "inner" side and an "outer" side, then an orientation is a choice of the inner or the outer side; the surface could also have "upward" and "downward" orientations

orthogonal vectors | vectors that form a right angle when placed in standard position

osculating circle | a circle that is tangent to a curve C at a point P and that shares the same curvature

osculating plane | the plane determined by the unit tangent and the unit normal vector

***p*-series** | a series of the form $\sum_{n=1}^{\infty} 1/n^p$

parallelogram method | a method for finding the sum of two vectors; position the vectors so they share the same initial point; the vectors then form two adjacent sides of a parallelogram; the sum of the vectors is the diagonal of that parallelogram

parameter | an independent variable that both x and y depend on in a parametric curve; usually represented by the variable t

parameter domain (parameter space) | the region of the uv-plane over which the parameters u and v vary for parameterization $\vec{\mathbf{r}}(u, v) = \langle x(u, v),\ y(u, v),\ z(u, v)\rangle$

parameterization of a curve | rewriting the equation of a curve defined by a function $y = f(x)$ as parametric equations

parameterized surface (parametric surface) | a surface given by a description of the form $\vec{\mathbf{r}}(u, v) = \langle x(u, v),\ y(u, v),\ z(u, v)\rangle$, where the parameters u and v vary over a parameter domain in the uv-plane

parametric curve | the graph of the parametric equations $x(t)$ and $y(t)$ over an interval $a \leq t \leq b$ combined with the equations

parametric equations | the equations $x = x(t)$ and $y = y(t)$ that define a parametric curve

parametric equations of a line | the set of equations $x = x_0 + ta, y = y_0 + tb$, and $z = z_0 + tc$ describing the line with direction vector $v = \langle a, b, c\rangle$ passing through point (x_0, y_0, z_0)

partial derivative | a derivative of a function of more than one independent variable in which all the variables but one are held constant

partial differential equation | an equation that involves an unknown function of more than one independent variable and one or more of its partial derivatives

partial fraction decomposition | a technique used to break down a rational function into the sum of simple rational functions

partial sum | the kth partial sum of the infinite series $\sum\limits_{n=1}^{\infty} a_n$ is the finite sum $S_k = \sum\limits_{n=1}^{k} a_n = a_1 + a_2 + a_3 + \cdots + a_k$

particular solution | member of a family of solutions to a differential equation that satisfies a particular initial condition

particular solution | a solution $y_p(x)$ of a differential equation that contains no arbitrary constants

partition | a set of points that divides an interval into subintervals

percentage error | the relative error expressed as a percentage

periodic function | a function is periodic if it has a repeating pattern as the values of x move from left to right

phase line | a visual representation of the behavior of solutions to an autonomous differential equation subject to various initial conditions

piecewise smooth curve | an oriented curve that is not smooth, but can be written as the union of finitely many smooth curves

piecewise-defined function | a function that is defined differently on different parts of its domain

planar transformation | a function T that transforms a region G in one plane into a region R in another plane by a change of variables

plane curve | the set of ordered pairs $(f(t), g(t))$ together with their defining parametric equations $x = f(t)$ and $y = g(t)$

point-slope equation | equation of a linear function indicating its slope and a point on the graph of the function

polar axis | the horizontal axis in the polar coordinate system corresponding to $r \geq 0$

polar coordinate system | a system for locating points in the plane. The coordinates are r, the radial coordinate, and θ, the angular coordinate

polar equation | an equation or function relating the radial coordinate to the angular coordinate in the polar coordinate system

polar rectangle | the region enclosed between the circles $r = a$ and $r = b$ and the angles $\theta = \alpha$ and $\theta = \beta$; it is described as $R = \{(r, \theta) \mid a \leq r \leq b,\ \alpha \leq \theta \leq \beta\}$

pole | the central point of the polar coordinate system, equivalent to the origin of a Cartesian system

polynomial function | a function of the form $f(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$

population growth rate | is the derivative of the population with respect to time

potential function | a scalar function f such that $\vec{\mathbf{\nabla}} f = \vec{\mathbf{F}}$

power function | a function of the form $f(x) = x^n$ for any positive integer $n \geq 1$

power law for limits | the limit law

$$\begin{gathered}\lim_{x \to a} (f(x))^n \\ = (\lim_{x \to a} f(x))^n \\ = L^n\end{gathered}$$

for every positive integer n

power reduction formula | a rule that allows an integral of a power of a trigonometric function to be exchanged for an integral involving a lower power

power rule | the derivative of a power function is a function in which the power on x becomes the coefficient of the term and the power on x in the derivative decreases by 1: If n is an integer, then $\frac{d}{dx}(x^n) = nx^{n-1}$

power series | a series of the form $\sum_{n=0}^{\infty} c_n x^n$ is a power series centered at $x = 0$; a series of the form $\sum_{n=0}^{\infty} c_n (x - a)^n$ is a power series centered at $x = a$

principal unit normal vector | a vector orthogonal to the unit tangent vector, given by the formula $\frac{\vec{\mathbf{T}}\ (t)}{\|\vec{\mathbf{T}}\ (t)\|}$

principal unit tangent vector | a unit vector tangent to a curve C

product law for limits | the limit law

$$\begin{gathered}\lim_{x \to a}(f(x) \cdot g(x)) \\ = \lim_{x \to a} f(x) \\ \cdot \lim_{x \to a} g(x) \\ = L \cdot M\end{gathered}$$

product rule | the derivative of a product of two functions is the derivative of the first function times the second function plus the derivative of the second function times the first function: $\frac{d}{dx}\big(f(x)g(x)\big) = f'(x)g(x) + g'(x)f(x)$

projectile motion | motion of an object with an initial velocity but no force acting on it other than gravity

propagated error | the error that results in a calculated quantity $f(x)$ resulting from a measurement error dx

quadratic function | a polynomial of degree 2; that is, a function of the form $f(x) = ax^2 + bx + c$ where $a \neq 0$

quadric surfaces | surfaces in three dimensions having the property that the traces of the surface are conic sections (ellipses, hyperbolas, and parabolas)

quotient law for limits | the limit law $\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{\lim_{x\to a} f(x)}{\lim_{x\to a} g(x)} = \frac{L}{M}$ for M≠0

quotient rule | the derivative of the quotient of two functions is the derivative of the first function times the second function minus the derivative of the second function times the first function, all divided by the square of the second function: $\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right) = \frac{f'(x)g(x) - g'(x)f(x)}{\big(g(x)\big)^2}$

radial coordinate | r the coordinate in the polar coordinate system that measures the distance from a point in the plane to the pole

radial field | a vector field in which all vectors either point directly toward or directly away from the origin; the magnitude of any vector depends only on its distance from the origin

radians | for a circular arc of length s on a circle of radius 1, the radian measure of the associated angle θ is s

radius of convergence | if there exists a real number $R > 0$ such that a power series centered at $x = a$ converges for $|x - a| < R$ and diverges for $|x - a| > R$, then R is the radius of convergence; if the power series only converges at $x = a$, the radius of convergence is $R = 0$; if the power series converges for all real numbers x, the radius of convergence is $R = \infty$

radius of curvature | the reciprocal of the curvature

radius of gyration | the distance from an object's center of mass to its axis of rotation

range | the set of outputs for a function

ratio test | for a series $\sum_{n=1}^{\infty} a_n$ with nonzero terms, let $\rho = \lim_{n\to\infty} |a_{n+1}/a_n|$; if $0 \leq \rho < 1$, the series converges absolutely; if $\rho > 1$, the series diverges; if $\rho = 1$, the test is inconclusive

rational function | a function of the form $f(x) = p(x)/q(x)$, where $p(x)$ and $q(x)$ are polynomials

recurrence relation | a recurrence relation is a relationship in which a term a_n in a sequence is defined in terms of earlier terms in the sequence

region | an open, connected, nonempty subset of $\mathbb{R}^2$

regular parameterization | parameterization $\mathbf{r}(u, v) = \langle x(u, v),\ y(u, v),\ z(u, v)\rangle$ such that $r_u \times r_v$ is not zero for point (u, v) in the parameter domain

regular partition | a partition in which the subintervals all have the same width

related rates | are rates of change associated with two or more related quantities that are changing over time

relative error | given an absolute error Δq for a particular quantity, $\frac{\Delta q}{q}$ is the relative error.

relative error | error as a percentage of the actual value, given by

$$\text{relative error} = \left|\frac{A - B}{A}\right| \cdot 100\%$$

remainder estimate | for a series $\sum_{n=}^{\infty} 1a_n$ with positive terms a_n and a continuous, decreasing function f such that $f(n) = a_n$ for all positive integers n, the remainder $R_N = \sum_{n=1}^{\infty} a_n - \sum_{n=1}^{N} a_n$ satisfies the following estimate:

$$\int_{N+1}^{\infty} f(x)\,dx < R_N < \int_N^{\infty} f(x)\,dx$$

removable discontinuity | A removable discontinuity occurs at a point a if $f(x)$ is discontinuous at a, but $\lim\limits_{x\to a} f(x)$ exists

reparameterization | an alternative parameterization of a given vector-valued function

restricted domain | a subset of the domain of a function f

riemann sum | an estimate of the area under the curve of the form $A \approx \sum_{i=1}^{n} f(x_i^*)\Delta x$

right-endpoint approximation | the right-endpoint approximation is an approximation of the area of the rectangles under a curve using the right endpoint of each subinterval to construct the vertical sides of each rectangle

right-hand rule | a common way to define the orientation of the three-dimensional coordinate system; when the right hand is curved around the z-axis in such a way that the fingers curl from the positive x-axis to the positive y-axis, the thumb points in the direction of the positive z-axis

***RLC* series circuit** | a complete electrical path consisting of a resistor, an inductor, and a capacitor; a second-order, constant-coefficient differential equation can be used to model the charge on the capacitor in an *RLC* series circuit

rolle's theorem | if f is continuous over $[a, b]$ and differentiable over (a, b), and if $f(a) = f(b)$, then there exists $c \in (a, b)$ such that $f'(c) = 0$

root function | a function of the form $f(x) = x^{1/n}$ for any integer $n \geq 2$

root law for limits | the limit law $\lim_{x\to a} \sqrt[n]{f(x)} = \sqrt[n]{\lim_{x\to a} f(x)} = \sqrt[n]{L}$ for all L if n is odd and for $L \geq 0$ if n is even

root test | for a series $\sum_{n=1}^{\infty} a_n$, let $\rho = \lim_{n\to\infty} \sqrt[n]{|a_n|}$; if $0 \leq \rho < 1$, the series converges absolutely; if $\rho > 1$, the series diverges; if $\rho = 1$, the test is inconclusive

rose | graph of the polar equation $r = a\cos 2\theta$ or $r = a\sin 2\theta$for a positive constant a

rotational field | a vector field in which the vector at point (x, y) is tangent to a circle with radius $r = \sqrt{x^2 + y^2}$; in a rotational field, all vectors flow either clockwise or counterclockwise, and the magnitude of a vector depends only on its distance from the origin

rulings | parallel lines that make up a cylindrical surface

saddle point | given the function $z = f(x, y)$, the point $(x_0, y_0, f(x_0, y_0))$ is a saddle point if both $f_x(x_0, y_0) = 0$ and $f_y(x_0, y_0) = 0$, but f does not have a local extremum at (x_0, y_0)

scalar | a real number

scalar equation of a plane | the equation $a(x - x_0) + b(y - y_0) + c(z - z_0) = 0$ used to describe a plane containing point $P = (x_0, y_0, z_0)$ with normal vector $n = \langle a, b, c\rangle$ or its alternate form $ax + by + cz + d = 0$, where $d = -ax_0 - by_0 - cz_0$

scalar line integral | the scalar line integral of a function f along a curve C with respect to arc length is the integral $\int_C f\ ds$, it is the integral of a scalar function f along a curve in a plane or in space; such an integral is defined in terms of a Riemann sum, as is a single-variable integral

scalar multiplication | a vector operation that defines the product of a scalar and a vector

scalar projection | the magnitude of the vector projection of a vector

secant | A secant line to a function $f(x)$ at a is a line through the point $(a, f(a))$ and another point on the function; the slope of the secant line is given by $m_{sec} = \frac{f(x) - f(a)}{x - a}$

second derivative test | suppose $f'(c) = 0$ and f'' is continuous over an interval containing c; if $f''(c) > 0$, then f has a local minimum at c; if $f''(c) < 0$, then f has a local maximum at c; if $f''(c) = 0$, then the test is inconclusive

separable differential equation | any equation that can be written in the form $y' = f(x)g(y)$

separation of variables | a method used to solve a separable differential equation

sequence | an ordered list of numbers of the form $a_1, a_2, a_3, \ldots$ is a sequence

sigma notation | (also, summation notation) the Greek letter sigma (Σ) indicates addition of the values; the values of the index above and below the sigma indicate where to begin the summation and where to end it

simple curve | a curve that does not cross itself

simple harmonic motion | motion described by the equation $x(t) = c_1\cos(\omega t) + c_2\sin(\omega t)$, as exhibited by an undamped spring-mass system in which the mass continues to oscillate indefinitely

simply connected region | a region that is connected and has the property that any closed curve that lies entirely inside the region encompasses points that are entirely inside the region

Simpson's rule | a rule that approximates $\int_a^b f(x)\,dx$ using the area under a piecewise quadratic function. The approximation S_n to $\int_a^b f(x)\,dx$ is given by

$$S_n = \frac{\Delta x}{3}\Big(f(x_0) + 4\,f(x_1) + 2\,f(x_2) + 4\,f(x_3) + 2\,f(x_4) + \cdots + 2\,f(x_{n-2}) + 4\,f(x_{n-1}) + f(x_n)\Big).$$

skew lines | two lines that are not parallel but do not intersect

slicing method | a method of calculating the volume of a solid that involves cutting the solid into pieces, estimating the volume of each piece, then adding these estimates to arrive at an estimate of the total volume; as the number of slices goes to infinity, this estimate becomes an integral that gives the exact value of the volume

slope | the change in y for each unit change in x

slope-intercept form | equation of a linear function indicating its slope and y-intercept

smooth | curves where the vector-valued function $\vec{r}(t)$ is differentiable with a non-zero derivative

solid of revolution | a solid generated by revolving a region in a plane around a line in that plane

solution curve | a curve graphed in a direction field that corresponds to the solution to the initial-value problem passing through a given point in the direction field

solution to a differential equation | a function $y = f(x)$ that satisfies a given differential equation

space curve | the set of ordered triples $(f(t), g(t), h(t))$ together with their defining parametric equations $x = f(t)$, $y = g(t)$ and $z = h(t)$

space-filling curve | a curve that completely occupies a two-dimensional subset of the real plane

speed | is the absolute value of velocity, that is, $|v(t)|$ is the speed of an object at time t whose velocity is given by $v(t)$

sphere | the set of all points equidistant from a given point known as the *center*

spherical coordinate system | a way to describe a location in space with an ordered triple (ρ, θ, φ), where ρ is the distance between P and the origin $(\rho \neq 0)$, θ is the same angle used to describe the location in cylindrical coordinates, and φ is the angle formed by the positive z-axis and line segment $\overline{OP}$, where O is the origin and $0 \leq \varphi \leq \pi$

squeeze theorem | states that if $f(x) \leq g(x) \leq h(x)$ for all $x \neq a$ over an open interval containing a and $\lim_{x\to a} f(x) = L = \lim_{x\to a} h(x)$ where L is a real number, then $\lim_{x\to a} g(x) = L$

standard equation of a sphere | $(x-a)^2 + (y-b)^2 + (z-c)^2 = r^2$ describes a sphere with center (a, b, c) and radius r

standard form | the form of a first-order linear differential equation obtained by writing the differential equation in the form $y' + p(x)y = q(x)$

standard form | an equation of a conic section showing its properties, such as location of the vertex or lengths of major and minor axes

standard unit vectors | unit vectors along the coordinate axes: $\hat{\imath} = \langle 1, 0\rangle$, $\hat{\jmath} = \langle 0, 1\rangle$

standard-position vector | a vector with initial point $(0, 0)$

steady-state solution | a solution to a nonhomogeneous differential equation related to the forcing function; in the long term, the solution approaches the steady-state solution

step size | the increment hh that is added to the xx value at each step in Euler's Method

Stokes' theorem | relates the flux integral over a surface S to a line integral around the boundary C of the surface S

stream function | if $\vec{F} = \langle P, Q\rangle$ is a source-free vector field, then stream function g is a function such that $P = g_y$ and $Q = -g_x$

sum law for limits | The limit law $\lim_{x\to a}(f(x) + g(x)) = \lim_{x\to a} f(x) + \lim_{x\to a} g(x) = L + M$

sum rule | the derivative of the sum of a function f and a function g is the same as the sum of the derivative of f and the derivative of g: $\frac{d}{dx}\big(f(x) + g(x)\big) = f'(x) + g'(x)$

surface | the graph of a function of two variables, $z = f(x, y)$

surface area | the surface area of a solid is the total area of the outer layer of the object; for objects such as cubes or bricks, the surface area of the object is the sum of the areas of all of its faces

surface area | the area of surface S given by the surface integral

$$\iint_S dS$$

surface independent | flux integrals of curl vector fields are surface independent if their evaluation does not depend on the surface but only on the boundary of the surface

surface integral | an integral of a function over a surface

surface integral of a scalar-valued function | a surface integral in which the integrand is a scalar function

surface integral of a vector field | a surface integral in which the integrand is a vector field

symmetric equations of a line | the equations $\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}$ describing the line with direction vector $v = \langle a, b, c\rangle$ passing through point (x_0, y_0, z_0)

symmetry about the

symmetry about the origin | the graph of a function f is symmetric about the origin if $(-x, -y)$ is on the graph of f whenever (x, y) is on the graph

symmetry principle | the symmetry principle states that if a region R is symmetric about a line I, then the centroid of R lies on I

table of values | a table containing a list of inputs and their corresponding outputs

tangent | A tangent line to the graph of a function at a point $(a, f(a))$ is the line that secant lines through $(a, f(a))$ approach as they are taken through points on the function with x-values that approach a; the slope of the tangent line to a graph at a measures the rate of change of the function at a

tangent line approximation (linearization) | since the linear approximation of f at $x = a$ is defined using the equation of the tangent line, the linear approximation of f at $x = a$ is also known as the tangent line approximation to f at $x = a$

tangent plane | given a function $f(x, y)$ that is differentiable at a point (x_0, y_0), the equation of the tangent plane to the surface $z = f(x, y)$ is given by $z = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$

tangent vector | to $\vec{r}(t)$ at $t = t_0$ any vector $\vec{v}$ such that, when the tail of the vector is placed at point $\vec{r}(t_0)$ on the graph, vector $\vec{v}$ is tangent to curve C

tangential component of acceleration | the coefficient of the unit tangent vector $\vec{T}$ when the acceleration vector is written as a linear combination of $\vec{T}$ and $\vec{N}$

Taylor polynomials | the n^{th}-degree Taylor polynomial for f at $x = a$ is $p_n(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n$

Taylor series | a power series at a that converges to a function f on some open interval containing a.

Taylor's theorem with remainder | for a function f and the n^{th}-degree Taylor polynomial for f at $x = a$, the remainder $R_n(x) = f(x) - p_n(x)$ satisfies $R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}$ for some c between x and a; if there exists an interval I containing a and a real number M such that $|\,f^{(n+1)}(x)\,| \leq M$ for all x in I, then $|R_n(x)| \leq \frac{M}{(n+1)!}|x-a|^{n+1}$

telescoping series | a telescoping series is one in which most of the terms cancel in each of the partial sums

term | the number a_n in the sequence a_n is called the nth term of the sequence

term-by-term differentiation of a power series | a technique for evaluating the derivative of a power series $\sum_{n=0}^{\infty} c_n(x-a)^n$ by evaluating the derivative of each term separately to create the new power series $\sum_{n=1}^{\infty} nc_n(x-a)^{n-1}$

term-by-term integration of a power series | a technique for integrating a power series $\sum_{n=0}^{\infty} c_n(x-a)^n$ by integrating each term separately to create the new power series $C + \sum_{n=0}^{\infty} c_n \frac{(x-a)^{n+1}}{n+1}$

terminal point | the endpoint of a vector

theorem of Pappus for volume | this theorem states that the volume of a solid of revolution formed by revolving a region around an external axis is equal to the area of the region multiplied by the distance traveled by the centroid of the region

three-dimensional rectangular coordinate system | a coordinate system defined by three lines that intersect at right angles; every point in space is described by an ordered triple (x, y, z) that plots its location relative to the defining axes

threshold population | the minimum population that is necessary for a species to survive

total area | total area between a function and the x-axis is calculated by adding the area above the x-axis and the area below the x-axis; the result is the same as the definite integral of the absolute value of the function

total differential | the total differential of the function $f(x, y)$ at (x_0, y_0) is given by the formula $dz = f_x(x_0, y_0)dx + fy(x_0, y_0)dy$

trace | the intersection of a three-dimensional surface with a coordinate plane

transcendental function | a function that cannot be expressed by a combination of basic arithmetic operations

transformation | a function that transforms a region GG in one plane into a region RR in another plane by a change of variables

transformation of a function | a shift, scaling, or reflection of a function

trapezoidal rule | a rule that approximates $\int_a^b f(x)\,dx$ using the area of trapezoids. The approximation T_n to $\int_a^b f(x)\,dx$ is given by

$$T_n = \frac{\Delta x}{2}\Big(f(x_0) + 2\,f(x_1) + 2\,f(x_2) + \cdots + 2\,f(x_{n-1}) + f(x_n)\Big).$$

tree diagram | illustrates and derives formulas for the generalized chain rule, in which each independent variable is accounted for

triangle inequality | If a and b are any real numbers, then $|a+b| \leq |a| + |b|$

triangle inequality | the length of any side of a triangle is less than the sum of the lengths of the other two sides

triangle method | a method for finding the sum of two vectors; position the vectors so the terminal point of one vector is the initial point of the other; these vectors then form two sides of a triangle; the sum of the vectors is the vector that forms the third side; the initial point of the sum is the initial point of the first vector; the terminal point of the sum is the terminal point of the second vector

trigonometric functions | functions of an angle defined as ratios of the lengths of the sides of a right triangle

trigonometric identity | an equation involving trigonometric functions that is true for all angles θ for which the functions in the equation are defined

trigonometric integral | an integral involving powers and products of trigonometric functions

trigonometric substitution | an integration technique that converts an algebraic integral containing expressions of the form $\sqrt{a^2-x^2}$, $\sqrt{a^2+x^2}$, or $\sqrt{x^2-a^2}$ into a trigonometric integral

triple integral | the triple integral of a continuous function $f(x,y,z)$ over a rectangular solid box B is the limit of a Riemann sum for a function of three variables, if this limit exists

triple integral in cylindrical coordinates | the limit of a triple Riemann sum, provided the following limit exists:

$$lim_{l,m,n\to\infty} \sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=1}^{n} f(r^*_{ijk}, \theta^*_{ijk}, s^*_{ijk}) r^*_{ijk} \Delta r \Delta\theta \Delta z$$

triple integral in spherical coordinates | the limit of a triple Riemann sum, provided the following limit exists:

$$lim_{l,m,n\to\infty} \sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=1}^{n} f(\rho^*_{ijk}, \theta^*_{ijk}, \varphi^*_{ijk})(\rho^*_{ijk})^2 \sin\varphi \Delta\rho \Delta\theta \Delta\varphi$$

Type I | a region D in the xy- plane is Type I if it lies between two vertical lines and the graphs of two continuous functions $g_1(x)$ and $g_2(x)$

Type II | a region D in the xy-plane is Type II if it lies between two horizontal lines and the graphs of two continuous functions $h_1(y)$ and $h_2(h)$

unbounded sequence | a sequence that is not bounded is called unbounded

unit vector | a vector with magnitude 1

unit vector field | a vector field in which the magnitude of every vector is 1

upper sum | a sum obtained by using the maximum value of $f(x)$ on each subinterval

variable of integration | indicates which variable you are integrating with respect to; if it is x, then the function in the integrand is followed by dx

vector | a mathematical object that has both magnitude and direction

vector addition | a vector operation that defines the sum of two vectors

vector difference | the vector difference $\vec{\mathbf{v}} - \vec{\mathbf{w}}$ is defined as $\vec{\mathbf{v}} + (-\vec{\mathbf{w}}) = \vec{\mathbf{v}} + (-1)\vec{\mathbf{w}}$

vector equation of a line | the equation $\vec{\mathbf{r}} = \vec{\mathbf{r}}_0 + t\vec{\mathbf{v}}$ used to describe a line with direction vector $\vec{\mathbf{v}} = \langle a,b,c\rangle$ passing through point $P = (x_0, y_0, z_0)$, where $\vec{\mathbf{r}}_0 = \langle x_0, y_0, z_0\rangle$, is the position vector of point P

vector equation of a plane | the equation $\vec{\mathbf{n}} \cdot \vec{PQ} = 0$, where P is a given point in the plane, Q is any point in the plane, and $\vec{\mathbf{n}}$ is a normal vector of the plane

vector field | measured in $\mathbb{R}^2$, an assignment of a vector $\vec{\mathbf{F}}(x,y)$ to each point (x,y) of a subset D of $\mathbb{R}^2$; in $\mathbb{R}^3$, an assignment of a vector $\vec{\mathbf{F}}(x,y,z)$ to each point (x,y,z) of a subset D of $\mathbb{R}^3$

vector line integral | the vector line integral of vector field $\vec{\mathbf{F}}$ along curve C is the integral of the dot product of $\vec{\mathbf{F}}$ with unit tangent vector $\vec{\mathbf{T}}$ of C with respect to arc length, $\int_C \vec{\mathbf{F}} \cdot \vec{\mathbf{T}}\,ds$; such an integral is defined in terms of a Riemann sum, similar to a single-variable integral

vector parameterization | any representation of a plane or space curve using a vector-valued function

vector projection | the component of a vector that follows a given direction

vector sum | the sum of two vectors, $\vec{\mathbf{v}}$ and $\vec{\mathbf{w}}$, can be constructed graphically by placing the initial point of $\vec{\mathbf{w}}$ at the terminal point of $\vec{\mathbf{v}}$; then the vector sum $\vec{\mathbf{v}} + \vec{\mathbf{w}}$ is the vector with an initial point that coincides with the initial point of $\vec{\mathbf{v}}$, and with a terminal point that coincides with the terminal point of $\vec{\mathbf{w}}$

vector-valued function | a function of the form $\vec{\mathbf{r}}(t) = f(t)\hat{\mathbf{i}} + g(t)\hat{\mathbf{j}}$ or $\vec{\mathbf{r}}(t) = f(t)\hat{\mathbf{i}} + g(t)\hat{\mathbf{j}} + h(t)\hat{\mathbf{k}}$,where the component functions f, g, and h are real-valued functions of the parameter t.

velocity vector | the derivative of the position vector

vertex | a vertex is an extreme point on a conic section; a parabola has one vertex at its turning point. An ellipse has two vertices, one at each end of the major axis; a hyperbola has two vertices, one at the turning point of each branch

vertical asymptote | A function has a vertical asymptote at $x=a$ if the limit as x approaches a from the right or left is infinite

vertical line test | given the graph of a function, every vertical line intersects the graph, at most, once

vertical trace | the set of ordered triples (c,y,z) that solves the equation $f(c,y)=z$ for a given constant $x=c$ or the set of ordered triples (x,d,z) that solves the equation $f(x,d)=z$ for a given constant $y=d$

washer method | a special case of the slicing method used with solids of revolution when the slices are washers

work | the amount of energy it takes to move an object; in physics, when a force is constant, work is expressed as the product of force and distance

work done by a force | work is generally thought of as the amount of energy it takes to move an object; if we represent an applied force by a vector $\vec{\mathbf{F}}$ and the displacement of an object by a vector $\vec{\mathbf{s}}$, then the work done by the force is the dot product of $\vec{\mathbf{F}}$ and $\vec{\mathbf{s}}$.

zero vector | the vector with both initial point and terminal point $(0,0)$

zeros of a function | when a real number x is a zero of a function f, $f(x)=0$

Detailed Licensing

Overview

Title: Calculus 3e (Apex)

Webpages: 127

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY-NC 3.0: 89% (113 pages)
- Undeclared: 11% (14 pages)

By Page

- Calculus 3e (Apex) - *CC BY-NC 3.0*
  - Front Matter - *Undeclared*
    - TitlePage - *Undeclared*
    - InfoPage - *Undeclared*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 1: Limits - *CC BY-NC 3.0*
    - 1.1: An Introduction to Limits - *CC BY-NC 3.0*
    - 1.2: Epsilon-Delta Definition of a Limit - *CC BY-NC 3.0*
    - 1.3: Finding Limits Analytically - *CC BY-NC 3.0*
    - 1.4: One Sided Limits - *CC BY-NC 3.0*
    - 1.5: Continuity - *CC BY-NC 3.0*
    - 1.6: Limits Involving Infinity - *CC BY-NC 3.0*
    - 1.E: Applications of Limits (Exercises) - *CC BY-NC 3.0*
  - 2: Derivatives - *CC BY-NC 3.0*
    - 2.1: Instantaneous Rates of Change- The Derivative - *CC BY-NC 3.0*
    - 2.2: Interpretations of the Derivative - *CC BY-NC 3.0*
    - 2.3: Basic Differentiation Rules - *CC BY-NC 3.0*
    - 2.4: The Product and Quotient Rules - *CC BY-NC 3.0*
    - 2.5: The Chain Rule - *CC BY-NC 3.0*
    - 2.6: Implicit Differentiation - *CC BY-NC 3.0*
    - 2.7: Derivatives of Inverse Functions - *CC BY-NC 3.0*
    - 2.E: Applications of Derivatives(Exercises) - *CC BY-NC 3.0*
  - 3: The Graphical Behavior of Functions - *CC BY-NC 3.0*
    - 3.1: Extreme Values - *CC BY-NC 3.0*
    - 3.2: The Mean Value Theorem - *CC BY-NC 3.0*
    - 3.3: Increasing and Decreasing Functions - *CC BY-NC 3.0*
    - 3.4: Concavity and the Second Derivative - *CC BY-NC 3.0*
    - 3.5: Curve Sketching - *CC BY-NC 3.0*
    - 3.E: Applications of the Graphical Behavior of Functions(Exercises) - *CC BY-NC 3.0*
  - 4: Applications of the Derivative - *CC BY-NC 3.0*
    - 4.1: Newton's Method - *CC BY-NC 3.0*
    - 4.2: Related Rates - *CC BY-NC 3.0*
    - 4.3: Optimization - *CC BY-NC 3.0*
    - 4.4: Differentials - *CC BY-NC 3.0*
    - 4.E: Applications of Derivatives (Exercises) - *CC BY-NC 3.0*
  - 5: Integration - *CC BY-NC 3.0*
    - 5.1: Antiderivatives and Indefinite Integration - *CC BY-NC 3.0*
    - 5.2: The Definite Integral - *CC BY-NC 3.0*
    - 5.3: Riemann Sums - *CC BY-NC 3.0*
    - 5.4: The Fundamental Theorem of Calculus - *CC BY-NC 3.0*
    - 5.5: Numerical Integration - *CC BY-NC 3.0*
    - 5.E: Applications of Integration (Exercises) - *CC BY-NC 3.0*
  - 6: Techniques of Integration - *CC BY-NC 3.0*
    - Front Matter - *Undeclared*
      - TitlePage - *Undeclared*
      - InfoPage - *Undeclared*
    - 6.1: Substitution - *CC BY-NC 3.0*
    - 6.2: Integration by Parts - *CC BY-NC 3.0*
    - 6.3: Trigonometric Integrals - *CC BY-NC 3.0*
    - 6.4: Trigonometric Substitution - *CC BY-NC 3.0*
    - 6.5: Partial Fraction Decomposition - *CC BY-NC 3.0*
    - 6.6: Hyperbolic Functions - *CC BY-NC 3.0*
    - 6.7: L'Hopital's Rule - *CC BY-NC 3.0*
    - 6.8: Improper Integration - *CC BY-NC 3.0*
    - 6.E: Applications of Antidifferentiation (Exercises) - *CC BY-NC 3.0*
    - Back Matter - *Undeclared*
      - Index - *Undeclared*
  - 7: Applications of Integration - *CC BY-NC 3.0*
    - 7.1: Area Between Curves - *CC BY-NC 3.0*

  - 7.2: Volume by Cross-Sectional Area- Disk and Washer Methods - *CC BY-NC 3.0*
  - 7.3: The Shell Method - *CC BY-NC 3.0*
  - 7.4: Arc Length and Surface Area - *CC BY-NC 3.0*
  - 7.5: Work - *CC BY-NC 3.0*
  - 7.6: Fluid Forces - *CC BY-NC 3.0*
  - 7.E: Applications of Integration (Exercises) - *CC BY-NC 3.0*
- 8: Sequences and Series - *CC BY-NC 3.0*
  - 8.1: Sequences - *CC BY-NC 3.0*
  - 8.2: Infinite Series - *CC BY-NC 3.0*
  - 8.3: Integral and Comparison Tests - *CC BY-NC 3.0*
  - 8.4: Ratio and Root Tests - *CC BY-NC 3.0*
  - 8.5: Alternating Series and Absolute Convergence - *CC BY-NC 3.0*
  - 8.6: Power Series - *CC BY-NC 3.0*
  - 8.7: Taylor Polynomials - *CC BY-NC 3.0*
  - 8.8: Taylor Series - *CC BY-NC 3.0*
  - 8.E: Applications of Sequences and Series (Exercises) - *CC BY-NC 3.0*
- 9: Curves in the Plane - *CC BY-NC 3.0*
  - 9.1: Conic Sections - *CC BY-NC 3.0*
  - 9.2: Parametric Equations - *CC BY-NC 3.0*
  - 9.3: Calculus and Parametric Equations - *CC BY-NC 3.0*
  - 9.4: Introduction to Polar Coordinates - *CC BY-NC 3.0*
  - 9.5: Calculus and Polar Functions - *CC BY-NC 3.0*
  - 9.E: Applications of Curves in a Plane (Exercises) - *CC BY-NC 3.0*
- 10: Vectors - *CC BY-NC 3.0*
  - 10.1: Introduction to Cartesian Coordinates in Space - *CC BY-NC 3.0*
  - 10.2: An Introduction to Vectors - *CC BY-NC 3.0*
  - 10.3: The Dot Product - *CC BY-NC 3.0*
  - 10.4: The Cross Product - *CC BY-NC 3.0*
  - 10.5: Lines - *CC BY-NC 3.0*
  - 10.6: Planes - *CC BY-NC 3.0*
  - 10.E: Applications of Vectors (Exercises) - *CC BY-NC 3.0*
- 11: Vector-Valued Functions - *CC BY-NC 3.0*
  - 11.1: Vector–Valued Functions - *CC BY-NC 3.0*
  - 11.2: Calculus and Vector-Valued Functions - *CC BY-NC 3.0*
  - 11.3: The Calculus of Motion - *CC BY-NC 3.0*
  - 11.4: Unit Tangent and Normal Vectors - *CC BY-NC 3.0*
  - 11.5: The Arc Length Parameter and Curvature - *CC BY-NC 3.0*
  - 11.E: Applications of Vector Valued Functions (Exercises) - *CC BY-NC 3.0*
- 12: Functions of Several Variables - *CC BY-NC 3.0*
  - 12.1: Introduction to Multivariable Functions - *CC BY-NC 3.0*
  - 12.2: Limits and Continuity of Multivariable Functions - *CC BY-NC 3.0*
  - 12.3: Partial Derivatives - *CC BY-NC 3.0*
  - 12.4: Differentiability and the Total Differential - *CC BY-NC 3.0*
  - 12.5: The Multivariable Chain Rule - *CC BY-NC 3.0*
  - 12.6: Directional Derivatives - *CC BY-NC 3.0*
  - 12.7: Tangent Lines, Normal Lines, and Tangent Planes - *CC BY-NC 3.0*
  - 12.8: Extreme Values - *CC BY-NC 3.0*
  - 12.E: Applications of Functions of Several Variables (Exercises) - *CC BY-NC 3.0*
- 13: Multiple Integration - *CC BY-NC 3.0*
  - 13.1: Iterated Integrals and Area - *CC BY-NC 3.0*
  - 13.2: Double Integration and Volume - *CC BY-NC 3.0*
  - 13.3: Double Integration with Polar Coordinates - *CC BY-NC 3.0*
  - 13.4: Center of Mass - *CC BY-NC 3.0*
  - 13.5: Surface Area - *CC BY-NC 3.0*
  - 13.6: Volume Between Surfaces and Triple Integration - *CC BY-NC 3.0*
  - 13.E: Applications of Multiple Integration (Exercises) - *CC BY-NC 3.0*
- 14: Appendix - *CC BY-NC 3.0*
  - 14.1: Section 1- - *CC BY-NC 3.0*
  - 14.2: Section 2- - *CC BY-NC 3.0*
  - 14.3: Section 3- - *CC BY-NC 3.0*
  - 14.4: Section 4- - *CC BY-NC 3.0*
  - 14.5: Section 5- - *CC BY-NC 3.0*
  - 14.6: Section 6- - *CC BY-NC 3.0*
- Back Matter - *Undeclared*
  - Index - *Undeclared*
  - Glossary - *Undeclared*
  - Detailed Licensing - *Undeclared*